PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:
HENRY TANG
BAKER BOTTS LLP
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112 0228

PCT

NOTIFICATION OF RECEIPT
OF DEMAND BY COMPETENT INTERNATIONAL
PRELIMINARY EXAMINING AUTHORITY

(PCT Rules 59.3(e) and 61.1(b), first sentence
and Administrative Instructions, Section 601(a))

Date of mailing *(day/month/year)*	06 OCT 2000
Applicant's or agent's file reference: 32312-PCT	**IMPORTANT NOTIFICATION**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US00/04505	22 FEB 00	19 FEB 99

Applicant
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF & NEW YORK

1. The applicant is hereby **notified** that this International Preliminary Examining Authority considers the following date as the date of receipt of the demand for international preliminary examination of the international application:

   05 SEP 2000

2. That date of receipt is:

   [x] the actual date of receipt of the demand by this Authority (Rule 61.1(b)).

   [ ] the actual date of receipt of the demand on behalf of this Authority (Rule 59.3(c)).

   [ ] the date on which this Authority has, in response to the invitation to correct defects in the demand (Form PCT/IPEA/404), received the required corrections.

3. [ ] **ATTENTION:** That date of receipt is **AFTER** the expiration of 19 months from the priority date. Consequently, the election(s) made in the demand does (do) not have the effect of postponing the entry into the national phase until 30 months from the priority date (or later in some Offices) (Article 39(1)). Therefore, the acts for entry into the national phase must be performed within 20 months from the priority date (or later in some Offices) (Article 22). For details, see the *PCT Applicant's Guide*, Volume II.

   [ ] *(If applicable)* This notification confirms the information given by telephone, facsimile transmission or in person on:

   _____

4. Only where paragraph 3 applies, a copy of this notification has been sent to the International Bureau.

On Docket For
8/19/01

Name and mailing address of the IPEA/	Authorized officer
Assistant Commissioner for Patent Box PCT Washington, D.C. 20231 Attn:RO/US Facsimile No. 703-305-3230	Telephone No.

Form PCT/IPEA/402 (July 1998)

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To: HENRY TANG
BAKER BOTTS LLP
20 ROCKEFELLER PLAZA
NEW YORK, NY 10112-0228

01 MAY -1 PM 12:36

PCT

NOTIFICATION OF TRANSMITTAL OF INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Rule 71.1)

Date of Mailing *(day/month/year)*	27 APR 2001
Applicant's or agent's file reference: 32312-PCT	IMPORTANT NOTIFICATION

International application No.	International filing date *(day/month/year)*	Priority Date *(day/month/year)*
PCT/US00/04505	22 FEBRUARY 2000	19 FEBRUARY 1999

Applicant
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices)(Article 39(1))(see also the reminder sent by the International Bureau with Form PCT/IB/301).

Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

ON DOCKET FOR 8/1/4/02

Name and mailing address of the IPEA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	STEPHEN HONG James R. Matthews
Facsimile No. (703) 305-3230	Telephone No. (703) 305-3900

Form PCT/IPEA/416 (July 1992)*

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 32312-PCT	**FOR FURTHER ACTION** See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/US00/04505	International filing date *(day/month/year)* 22 FEBRUARY 2000	Priority date *(day/month/year)* 19 FEBRUARY 1999
International Patent Classification (IPC) or national classification and IPC IPC(7): G06F 17/27 and US Cl.: 707/500, 501, 530; 704/9, 10		
Applicant THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 3 sheets.

   ☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority. (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 0 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of report with regard to novelty, inventive step or industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand 05 SEPTEMBER 2000	Date of completion of this report 07 APRIL 2001
Name and mailing address of the IPEA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHEN HONG James R. Matthews Telephone No. (703) 305-3900

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US00/04505

I. Basis of the report

1. With regard to the elements of the international application:*

[X] the international application as originally filed

[X] the description:
pages 1-15 , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of _____

[X] the claims:
pages 16-22 , as originally filed
pages NONE , as amended (together with any statement) under Article 19
pages NONE , filed with the demand
pages NONE , filed with the letter of _____

[X] the drawings:
pages 1, 3-7 , as originally filed
pages NONE , filed with the demand
pages Page 2, filed with the letter , filed with the letter of _____

[X] the sequence listing part of the description:
pages NONE , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of _____

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.
These elements were available or furnished to this Authority in the following language _____ which is:

[ ] the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).

[ ] the language of publication of the international application (under Rule 48.3(b)).

[ ] the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).

3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international

[ ] contained in the international application in printed form.

[ ] filed together with the international application in computer readable form.

[ ] furnished subsequently to this Authority in written form.

[ ] furnished subsequently to this Authority in computer readable form.

[ ] The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

[ ] The statement that the information recorded in computer readable form is identical to the writen sequence listing has been furnished.

4. [X] The amendments have resulted in the cancellation of:

[X] the description, pages NONE

[X] the claims, Nos. NONE

[X] the drawings, sheets/fig NONE

5. [ ] This report has been drawn as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

** Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17).*

*** Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.*

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US00/04505

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. statement

Novelty (N)	Claims	1-37	YES
	Claims	NONE	NO
Inventive Step (IS)	Claims	NONE	YES
	Claims	1-37	NO
Industrial Applicability (IA)	Claims	1-37	YES
	Claims	NONE	NO

2. citations and explanations (Rule 70.7)

Claims 1-37 lack an inventive step under PCT Article 33(3) as being obvious over Doi in view of Kupiec et al.

As per claims 1-37, Doi teaches the claimed feature of generating a summary by extracting sentences from the input document and parsing the extracted sentences into components (col.1, lines 52-60); sentence reduction processing which is performed to mark components which can be removed from the parsed sentences (FIG.4(B); col.3, lines 45-67); evaluating the importance of the context of the sentences and linguistic knowledge based processing (see the parts of speech analysis in FIG.3); combining sentences for identifying sentence combination operations and establishing rules for applying the sentence combination operations to merge at least two sentences and removing the unwanted portions of the sentences (col.5, line 20-35); and generating a summary of the document (see FIG.8; col.5, lines 35-42). However, Doi does not explicitly teach the use of the probablistic importance processing. Nevertheless, Kupiec et al. shows the probablistic processing for evaluating the importance in the summary generation system (col.1, lines 57 to col.2, line 17, "...the probability of observing a value of a particular feature in a sentence included in the summary and the probability of that feature taking each of its possible values..."). It would have been obvious to a person of ordinary skill in the art at the time of the invention to have incorporated Kupiec's probablistic processing into Doi, since Kupiec provided the motivation by pointing out that the probablistic model ensures more important parts of the sentences to be chosen for the summary.

Furthermore, although the prior art does not explicitly disclose the use of the "Hidden Markov Model" or "Viterbi algorithm" for the probability model, such were well known in the art, and thus, would have been obvious to a person of ordinary skill in the art at the time of the invention.

---------- NEW CITATIONS ----------

NONE

PATENT COOPERATION TREATY

BAKER BOTTS L.L.P.

00 AUG 22 AM 10: 33

From the INTERNATIONAL SEARCHING AUTHORITY

To: HENRY TANG
BAKER BOTTS LLP
20 ROCKEFELLER PLAZA
NEW YORK NY 10112-0228

PCT

TO

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

	Date of Mailing *(day/month/year)* 15 AUG 2000
Applicant's or agent's file reference 32312-PCT	**FOR FURTHER ACTION** **See paragraphs 1 and 4 below**
International application No. PCT/US00/04505	International filing date *(day/month/year)* 22 FEBRUARY 2000
Applicant THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK	

1. [X] The applicant is hereby notified that the international search report has been established and is transmitted herewith.

Filing of amendments and statement under Article 19:
The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

When? The time limit for filing such amendments is normally 2 months from the date of transmittal of the international search report; however, for more details, see the notes on the accompanying sheet.

Where? Directly to the International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland
Facsimile No.: (41-22) 740.14.35

Docketed
For 10 / 15 /2000 By

For more detailed instructions, see the notes on the accompanying sheet.

2. [ ] The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. [ ] **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

[ ] the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

[ ] no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Further action(s):** The applicant is reminded of the following:

Shortly after **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in rules 90 *bis* 1 and 90 *bis* 3, respectively, before the completion of the technical preparations for international publication.

Within **19 months** from the priority date, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later).

Within **20 months** from the priority date, the applicant must perform the prescribed acts for entry into the national phase before all designated Offices which have not been elected in the demand or in a later election within 19 months from the priority date or could not be elected because they are not bound by Chapter II.

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer STEPHEN HONG Eugenia Zagan
Facsimile No. (703) 305-3230	Telephone No. (703) 305-3900

Form PCT/ISA/220 (July 1998)★ *(See notes on accompanying sheet)*

X Copy of Search Report & References IL pocket #

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 32312-PCT	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
International application No. PCT/US00/04505	International filing date *(day/month/year)* 22 FEBRUARY 2000	(Earliest) Priority Date *(day/month/year)* 19 FEBRUARY 1999
Applicant THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.
   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).
   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing:
   [ ] contained in the international application in written form.
   [ ] filed together with the international application in computer readable form.
   [ ] furnished subsequently to this Authority in written form.
   [ ] furnished subsequently to this Authority in computer readable form.
   [ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
   [ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has b een furnished.
2. [ ] **Certain claims were found unsearchable** (See Box I).
3. [ ] **Unity of invention is lacking** (See Box II).
4. With regard to the **title**,
   [X] the text is approved as submitted by the applicant.
   [ ] the text has been established by this Authority to read as follows:
5. With regard to the **abstract**,
   [ ] the text is approved as submitted by the applicant.
   [X] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.
6. The figure of the **drawings** to be published with the abstract is Figure No. 1
   [ ] as suggested by the applicant.
   [X] because the applicant failed to suggest a figure.
   [ ] because this figure better characterizes the invention.
   [ ] None of the figures.

Box III TEXT OF THE ABSTRACT (Continuation of item 5 of the first sheet)

The technical features mentioned in the abstract do not include a reference sign between parentheses (PCT Rule 8.1(d)).

NEW ABSTRACT

A summary of an input document is generated by extracting at least one sentence from the document and parsing the extracted sentences into components, such as in a parse tree (110). Sentence reduction processing is performed to mark components which can be removed from the parse trees (135). Sentence reduction can include context importance processing, probabilistic processing, and linguistic knowledge based processing, probabilistic processing includes identifying sentence combination operations and establishing rules for applying the sentence combination operations to mark the parse trees to merge at least two sentences (140). Sentence combination processing also provides a paste operation to operate on the marked components to effect the indicated removal and combination of sentence components, thereby generating summary sentences for the input document.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04505

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :G06F 17/27
US CL : 707/500, 501, 530; 704/9, 10
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/500, 501, 530; 704/9, 10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST database
search terms: summary, summarization, document,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,778,397 A (KUPICE et al) 07 July 1998, col.3, line 37 to col.10, line 35	1-37
Y	US 5,077,668 A (DOI) 31 December 1991, col.2, line 50 to col.4, line 44.	1-37
A, P	US 5,918,240 A (KUPIEC et al) 29 June 1999, ALL	1-37
A	US 5,838,323 A (ROSE et al) 17 November 1998, ALL	1-37
A, P	US 5,924,108 A (FEIN et al) 13 July 1999, ALL	1-37

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
11 JULY 2000

Date of mailing of the international search report
15 AUG 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
STEPHEN HONG
Telephone No. (703) 305-3900

Form PCT/ISA/210 (second sheet) (July 1998) ★

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70) 14

Applicant's or agent's file reference 32312-PCT	**FOR FURTHER ACTION**	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/US00/04505	International filing date *(day/month/year)* 22 FEBRUARY 2000	Priority date *(day/month/year)* 19 FEBRUARY 1999
International Patent Classification (IPC) or national classification and IPC IPC(7): G06F 17/27 and US Cl.: 707/500, 501, 530; 704/9, 10		
Applicant THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 3 sheets.

   ☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority. (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 0 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of report with regard to novelty, inventive step or industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
05 SEPTEMBER 2000	07 APRIL 2001
Name and mailing address of the IPEA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHEN HONG James R. Matthews Telephone No. (703) 305-3900

Form PCT/IPEA/409 (cover sheet) (July 1998)★

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US00/04505

I. Basis of the report

1. With regard to the elements of the international application:*

[X] the international application as originally filed

[X] the description:
pages 1-15 , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of _____

[X] the claims:
pages 16-22 , as originally filed
pages NONE , as amended (together with any statement) under Article 19
pages NONE , filed with the demand
pages NONE , filed with the letter of _____

[X] the drawings:
pages 1, 3-7 , as originally filed
pages NONE , filed with the demand
pages Page 2, filed with the letter , filed with the letter of _____

[X] the sequence listing part of the description: NONE , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of _____

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.
These elements were available or furnished to this Authority in the following language _____ which is:

[ ] the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).

[ ] the language of publication of the international application (under Rule 48.3(b)).

[ ] the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).

3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international

[ ] contained in the international application in printed form.

[ ] filed together with the international application in computer readable form.

[ ] furnished subsequently to this Authority in written form.

[ ] furnished subsequently to this Authority in computer readable form.

[ ] The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

[ ] The statement that the information recorded in computer readable form is identical to the writen sequence listing has been furnished.

4. [X] The amendments have resulted in the cancellation of:

[X] the description, pages NONE

[X] the claims, Nos. NONE

[X] the drawings, sheets/~~fig~~ NONE

5. [ ] This report has been drawn as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

** Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17).*

*** Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.*

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US00/04505

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. statement

Novelty (N)	Claims	1-37	YES
	Claims	NONE	NO
Inventive Step (IS)	Claims	NONE	YES
	Claims	1-37	NO
Industrial Applicability (IA)	Claims	1-37	YES
	Claims	NONE	NO

2. citations and explanations (Rule 70.7)

Claims 1-37 lack an inventive step under PCT Article 33(3) as being obvious over Doi in view of Kupiec et al.

As per claims 1-37, Doi teaches the claimed feature of generating a summary by extracting sentences from the input document and parsing the extracted sentences into components (col.1, lines 52-60); sentence reduction processing which is performed to mark components which can be removed from the parsed sentences (FIG.4(B); col.3, lines 45-67); evaluating the importance of the context of the sentences and linguistic knowledge based processing (see the parts of speech analysis in FIG.3); combining sentences for identifying sentence combination operations and establishing rules for applying the sentence combination operations to merge at least two sentences and removing the unwanted portions of the sentences (col.5, line 20-35); and generating a summary of the document (see FIG.8; col.5, lines 35-42). However, Doi does not explicitly teach the use of the probablistic importance processing. Nevertheless, Kupiec et al. shows the probablistic processing for evaluating the importance in the summary generation system (col.1, lines 57 to col.2, line 17, "...the probability of observing a value of a particular feature in a sentence included in the summary and the probability of that feature taking each of its possible values..."). It would have been obvious to a person of ordinary skill in the art at the time of the invention to have incorporated Kupiec's probablistic processing into Doi, since Kupiec provided the motivation by pointing out that the probablistic model ensures more important parts of the sentences to be chosen for the summary.

Furthermore, although the prior art does not explicitly disclose the use of the "Hidden Markov Model" or "Viterbi algorithm" for the probability model, such were well known in the art, and thus, would have been obvious to a person of ordinary skill in the art at the time of the invention.

---------- NEW CITATIONS ----------

NONE

The demand must be filed directly with the competent International Preliminary Examining Authority or, if two or more Authorities are with the one chosen by the applicant. Give the full name or two-letter code of that Authority indicated by the applicant on the line

IPEA/ US

PCT

CHAPTER II

DEMAND

under Article 31 of the Patent Cooperation Treaty:
The undersigned requests that the international application specified below be the subject of international preliminary examination according to the Patent Cooperation Treaty and hereby elects all eligible States (except where otherwise indicated).

For International Preliminary Examining Authority use only

Identification of IPEA	Date of receipt of DEMAND

Box No. I IDENTIFICATION OF THE INTERNATIONAL APPLICATION		Applicant's or agent's file reference 32312-PCT
International application No. PCT/US00/04505	International filing date *(day/month/year)* 22 February 2000 (22.02.00)	(Earliest) Priority date *(day/month/year)* 19 February 1999 (19.02.99)

Title of invention
CUT AND PASTE DOCUMENT SUMMARIZATION SYSTEM AND METHOD

Box No. II APPLICANT(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*	Telephone No.:
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK 116th Street and Broadway New York, NY 10027 US	Facsimile No.:
	Teleprinter No.:

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

MCKEOWN, KATHLEEN R.
20 Prospect Road
Wayne, NJ 07470
US

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

JING, HONGYAN
521 West 112th Street, Apt. 73C
New York, NY 10025
US

State *(that is, country)* of nationality:	State *(that is, country)* of residence:
US	US

☐ Further applicants are indicated on a continuation sheet.

International application No.
PCT/US00/04505

Box No. III AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The following person is ☒ agent ☐ common representative

and ☒ has been appointed earlier and represents the applicant(s) also for international preliminary examination.

☐ is hereby appointed and any earlier appointment of (an) agent(s) /common representative is hereby revoked.

☐ is hereby appointed, specifically for the procedure before the International Preliminary Examining Authority, in addition to the agent(s)/common representative appointed earlier.

Name and address: *(Family name followed by given name; for a legal entity, full official The address must include postal code and name of country.)*

TANG, HENRY and
ACKERMAN, PAUL D.
Baker Botts LLP
30 Rockefeller Plaza
New York, NY 10112-0228
US

Telephone No.:
(212) 705-5000

Facsimile No.:
(212) 705-5020

Teleprinter No.:

☐ Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Box No. IV BASIS FOR INTERNATIONAL PRELIMINARY EXAMINATION

Statement concerning amendments:*

1. The applicant wishes the international preliminary examination **to start on the basis of:**

☒ the international application as originally filed.

the description
☐ as originally filed
☐ as amended under Article 34

the claims
☐ as originally filed
☐ as amended under Article 19 (together with any accompanying statement)
☐ as amended under Article 34

the drawings
☐ as originally filed
☐ as amended under Article 34

2. ☐ The applicant wishes any amendment to the claims under Article 19 to be considered as reversed.

3. ☐ The applicant wishes the start of the international preliminary examination **to be postponed** until the expiration of 20 months from the priority date unless the International Preliminary Examing Authority receives a copy of any amendments made under Article 19 or a notice from the applicant that he does not wish to make such amendments (Rule 69.1(d)). *(This check-box may be marked only where the time limit under Article 19 has not yet expired.)*

* Where no check-box is marked, international preliminary examination will start on the basis of the international application as originally filed or, where a copy of amendments to the claims under Article 19 and/or amendments of the international application under Article 34 are received by the International Preliminary Examining Authority before it has begun to draw up a written opinion or the international preliminary examination report, as so amended.

Language for the purposes of international preliminary examination: English

☒ which is the language in which the international application was filed.
☐ which is the language of a translation furnished for the purposes of international search.
☐ which is the language of publication of the international application.
☐ which is the language of the translation (to be) furnished for the purposes of international preliminary examination.

Box No. V ELECTION OF STATES

The applicant hereby **elects all eligible States** *(that is, all States which have been designated and which are bound by Chapter II of the PCT)*

excluding the following States which the applicant wishes **not to elect:**

International application No.
PCT/US00/04505

Box No. VI CHECK LIST

The demand is accompanied by the following elements, in the language referred to in Box No. IV, for the purposes of international preliminary examination:			For International Preliminary Examining Authority use only: received	For International Preliminary Examining Authority use only: not received
1. translation of international application	:	sheets	☐	☐
2. amendments under Article 34	:	sheets	☐	☐
3. copy (or where required, translation) of amendments under Article 19	:	sheets	☐	☐
4. copy (or, where required, translation) of statement under Article 19	:	sheets	☐	☐
5. letter	:	sheets	☐	☐
6. other *(specify)*	:	sheets	☐	☐

The demand is also accompanied by the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ separate signed power of attorney
3. ☐ copy of general power of attorney; reference number, if any:
4. ☐ statement explaining lack of signature
5. ☐ nucleotide and or amino acid sequence listing in computer readable form
6. ☒ other *(specify)*: Transmittal Letter

Box No. VII SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the demand).

Paul D. Ackerman (Agent)

For International Preliminary Examining Authority use only

1. Date of actual receipt of DEMAND:
2. Adjusted date of receipt of demand due to CORRECTIONS under Rule 60.1(b):
3. ☐ The date of receipt of the demand is AFTER the expiration of 19 months from the priority date and item 4 or 5, below, does not apply. ☐ The applicant has been informed accordingly.
4. ☐ The date of receipt of the demand is WITHIN the period of 19 months from the priority date as extended by virtue of Rule 80.5.
5. ☐ Although the date of receipt of the demand is after the expiration of 19 months from the priority date, the delay in arrival is EXCUSED pursuant to Rule 82.

For International Bureau use only

Demand received from IPEA on:

CHAPTER II

PCT

FEE CALCULATION SHEET

Annex to the Demand for international preliminary examination

International application No.	PCT/US00/04505
Applicant's or agent's file reference	32312-PCT

For International Preliminary Examining Authority use only

Date stamp of the IPEA

Applicant

THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Calculation of prescribed fees

1. Preliminary examination fee	490.00	P
2. Handling fee *(Applicants from certain States are entitled to a reduction of 75% of the handling fee. Where the applicant is (or all applicants are) so entitled, the amount to be entered at H is 25% of the handling fee.)*	153.00	H
3. Total of prescribed fees Add the amounts entered at P and H and enter total in the TOTAL box	643.00	TOTAL

Mode of Payment

- ☐ authorization to charge deposit account with the IPEA (see below)
- ☒ cheque
- ☐ postal money order
- ☐ bank draft
- ☐ cash
- ☐ revenue stamps
- ☐ coupons
- ☐ other *(specify)*:

Deposit Account Authorization *(this mode of payment may not be available at all IPEAs)*

The IPEA/ US ☐ is hereby authorized to charge the total fees indicated above to my deposit account.

☒ *(this check-box may be marked only if the conditions for deposit accounts of the IPEA so permit)* is hereby authorized to charge any deficiency or credit any overpayment in the total fees indicated above to my deposit account.

02-4377	5 September 2000	
Deposit Account Number	Date *(day/month/year)*	Signature

PATENT COOPERATION TREATY

32312 PCT

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF RECEIPT OF RECORD COPY

(PCT Rule 24.2(a))

To:

TANG, Henry
Baker & Botts LLP
30 Rockefeller Plaza
New York, NY 10112-0228
ETATS-UNIS D'AMERIQUE

BAKER BOTTS L.L.P
00 MAY 23 PM 12:17

TO

Date of mailing (day/month/year)	IMPORTANT NOTIFICATION
11 May 2000 (11.05.00)	
Applicant's or agent's file reference	**International application No.**
32312-PCT	PCT/US00/04505

The applicant is hereby notified that the International Bureau has received the record copy of the international application as detailed below.

Name(s) of the applicant(s) and State(s) for which they are applicants:

THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK (for all designated States except US)
MCKEOWN, Kathleen, R. et al (for US)

International filing date	:	22 February 2000 (22.02.00)
Priority date(s) claimed	:	19 February 1999 (19.02.99)
Date of receipt of the record copy by the International Bureau	:	26 April 2000 (26.04.00)
List of designated Offices	:	

AP :GH,GM,KE,LS,MW,SD,SL,SZ,TZ,UG,ZW
EA :AM,AZ,BY,KG,KZ,MD,RU,TJ,TM
EP :AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE
OA :BF,BJ,CF,CG,CI,CM,GA,GN,GW,ML,MR,NE,SN,TD,TG
National :AE,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,CA,CH,CN,CR,CU,CZ,DE,DK,DM,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

The receiving Office was closed for business on 21 February 2000 (21.02.00)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer:

Marie-José Devillard

Telephone No. (41-22) 338.83.38

Continuation of Form PCT/IB/301

NOTIFICATION OF RECEIPT OF RECORD COPY

Date of mailing (day/month/year)	IMPORTANT NOTIFICATION
11 May 2000 (11.05.00)	
Applicant's or agent's file reference	**International application No.**
32312-PCT	PCT/US00/04505

ATTENTION

The applicant should carefully check the data appearing in this Notification. In case of any discrepancy between these data and the indications in the international application, the applicant should immediately inform the International Bureau.

In addition, the applicant's attention is drawn to the information contained in the Annex, relating to:

- [X] time limits for entry into the national phase
- [ ] confirmation of precautionary designations
- [X] requirements regarding priority documents

A copy of this Notification is being sent to the receiving Office and to the International Searching Authority.

INFORMATION ON TIME LIMITS FOR ENTERING THE NATIONAL PHASE

The applicant is reminded that the "national phase" must be entered before each of the designated Offices indicated in the Notification of Receipt of Record Copy (Form PCT/IB/301) by paying national fees and furnishing translations, as prescribed by the applicable national laws.

The time limit for performing these procedural acts is **20 MONTHS** from the priority date or, for those designated States which the applicant elects in a demand for international preliminary examination or in a later election, **30 MONTHS** from the priority date, provided that the election is made before the expiration of 19 months from the priority date. Some designated (or elected) Offices have fixed time limits which expire even later than 20 or 30 months from the priority date. In other Offices an extension of time or grace period, in some cases upon payment of an additional fee, is available.

In addition to these procedural acts, the applicant may also have to comply with other special requirements applicable in certain Offices. **It is the applicant's responsibility** to ensure that the necessary steps to enter the national phase are taken in a timely fashion. Most designated Offices do not issue reminders to applicants in connection with the entry into the national phase.

For detailed information about the procedural acts to be performed to enter the national phase before each designated Office, the applicable time limits and possible extensions of time or grace periods, and any other requirements, see the relevant Chapters of Volume II of the PCT Applicant's Guide. Information about the requirements for filing a demand for international preliminary examination is set out in Chapter IX of Volume I of the PCT Applicant's Guide.

GR and ES became bound by PCT Chapter II on 7 September 1996 and 6 September 1997, respectively, and may, therefore, be elected in a demand or a later election filed on or after 7 September 1996 and 6 September 1997, respectively, regardless of the filing date of the international application. (See second paragraph above.)

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

CONFIRMATION OF PRECAUTIONARY DESIGNATIONS

This notification lists only specific designations made under Rule 4.9(a) in the request. It is important to check that these designations are correct. Errors in designations can be corrected where precautionary designations have been made under Rule 4.9(b). The applicant is hereby reminded that any precautionary designations may be confirmed according to Rule 4.9(c) before the expiration of 15 months from the priority date. If it is not confirmed, it will automatically be regarded as withdrawn by the applicant. There will be no reminder and no invitation. Confirmation of a designation consists of the filing of a notice specifying the designated State concerned (with an indication of the kind of protection or treatment desired) and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.

REQUIREMENTS REGARDING PRIORITY DOCUMENTS

For applicants who have not yet complied with the requirements regarding priority documents, the following is recalled.

Where the priority of an earlier national, regional or international application is claimed, the applicant must submit a copy of the said earlier application, certified by the authority with which it was filed ("the priority document") to the receiving Office (which will transmit it to the International Bureau) or directly to the International Bureau, before the expiration of 16 months from the priority date, provided that any such priority document may still be submitted to the International Bureau before that date of international publication of the international application, in which case that document will be considered to have been received by the International Bureau on the last day of the 16-month time limit (Rule 17.1(a)).

Where the priority document is issued by the receiving Office, the applicant may, instead of submitting the priority document, request the receiving Office to prepare and transmit the priority document to the International Bureau. Such request must be made before the expiration of the 16-month time limit and may be subjected by the receiving Office to the payment of a fee (Rule 17.1(b)).

If the priority document concerned is not submitted to the International Bureau or if the request to the receiving Office to prepare and transmit the priority document has not been made (and the corresponding fee, if any, paid) within the applicable time limit indicated under the preceding paragraphs, any designated State may disregard the priority claim, provided that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity to furnish the priority document within a time limit which is reasonable under the circumstances.

Where several priorities are claimed, the priority date to be considered for the purposes of computing the 16-month time limit is the filing date of the earliest application whose priority is claimed.

PCT/US00/04505
32312
PCT

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION CONCERNING SUBMISSION OR TRANSMITTAL OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

To:

BAKER BOTTS L.L.P.

00 JUN 30 AM 10:47

TANG, Henry
Baker & Botts LLP
30 Rockefeller Plaza
New York, NY 10112-0228
ETATS-UNIS D'AMERIQUE

Date of mailing (day/month/year) 15 June 2000 (15.06.00)	
Applicant's or agent's file reference 32312-PCT	**IMPORTANT NOTIFICATION**
International application No. PCT/US00/04505	International filing date (day/month/year) 22 February 2000 (22.02.00)
International publication date (day/month/year) Not yet published	Priority date (day/month/year) 19 February 1999 (19.02.99)
Applicant THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK et al	

1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).

2. This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.

3. An **asterisk(*)** appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

4. The **letters "NR"** appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

Priority date	Priority application No.	Country or regional Office or PCT receiving Office	Date of receipt of priority document
19 Febr 1999 (19.02.99)	60/120,657	US	15 May 2000 (15.05.00)

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer

Tessadel PAMPLIEGA

Telephone No. (41-22) 338.83.38

cn

PATENT COOPERATION TREATY

32312 RCT

From the INTERNATIONAL BUREAU

PCT

NOTICE INFORMING THE APPLICANT OF THE COMMUNICATION OF THE INTERNATIONAL APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

To:
TANG, Henry
Baker & Botts LLP
30 Rockefeller Plaza
New York, NY 10112-0228
ETATS-UNIS D'AMERIQUE

BAKER BOTTS L.L.P.
01 FEB -6 AM 11:17

Date of mailing (day/month/year)	
25 January 2001 (25.01.01)	
Applicant's or agent's file reference	**IMPORTANT NOTICE**
32312-PCT	

International application No.	International filing date (day/month/year)	Priority date (day/month/year)
PCT/US00/04505	22 February 2000 (22.02.00)	19 February 1999 (19.02.99)

Applicant
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK et al

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:
   AU,KP,KR,US

   In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:
   AE,AL,AM,AP,AT,AZ,BA,BB,BG,BR,BY,CA,CH,CN,CR,CU,CZ,DE,DK,DM,EA,EE,EP,ES,FI,GB,GD,
   GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,
   NO,NZ,OA,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZW
   The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on
   25 January 2001 (25.01.01) under No. WO 01/06408

REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, a **demand for international preliminary examination** must be filed with the competent International Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))

If the applicant wishes to proceed with the international application in the **national phase**, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

Final 8/19/01 Encl. Inquiry

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer J. Zahra
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

INFORMATION CONCERNING ELECTED OFFICES NOTIFIED OF THEIR ELECTION

(PCT Rule 61.3)

To:
TANG, Henry
Baker & Botts LLP
30 Rockefeller Plaza
New York, NY 10112-0228
ETATS-UNIS D'AMERIQUE

Date of mailing (day/month/year) 25 January 2001 (25.01.01)	
Applicant's or agent's file reference 32312-PCT	**IMPORTANT INFORMATION**

International application No.	**International filing date** (day/month/year)	**Priority date** (day/month/year)
PCT/US00/04505	22 February 2000 (22.02.00)	19 February 1999 (19.02.99)

Applicant
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK et al

1. The applicant is hereby informed that the International Bureau has, according to Article 31(7), notified each of the following Offices of its election:

   AP :GH,GM,KE,LS,MW,SD,SL,SZ,TZ,UG,ZW
   EP :AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE
   National :AU,BG,CA,CN,CZ,DE,IL,JP,KP,KR,MN,NO,NZ,PL,RO,RU,SE,SK,US

2. The following Offices have waived the requirement for the notification of their election; the notification will be sent to them by the International Bureau only upon their request:

   EA :AM,AZ,BY,KG,KZ,MD,RU,TJ,TM
   OA :BF,BJ,CF,CG,CI,CM,GA,GN,GW,ML,MR,NE,SN,TD,TG
   National :AE,AL,AM,AT,AZ,BA,BB,BR,BY,CH,CR,CU,DK,DM,EE,ES,FI,GB,GD,GE,GH, GM,HR,HU,ID,IN,IS,KE,KG,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MW,MX,PT,SD, SG,SI,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZW

3. The applicant is reminded that he must enter the "national phase" **before the expiration of 30 months from the priority date** before each of the Offices listed above. This must be done by paying the national fee(s) and furnishing , if prescribed, a translation of the international application (Article 39(1)(a)), as well as, where applicable, by furnishing a translation of any annexes of the international preliminary examination report (Article 36(3)(b) and Rule 74.1).

   Some offices have fixed time limits expiring later than the above-mentioned time limit. For detailed information about the applicable time limits and the acts to be performed upon entry into the national phase before a particular Office, see Volume II of the PCT Applicant's Guide.

   The entry into the European regional phase is postponed **until 31 months from the priority date** for all States designated for the purposes of obtaining a European patent.

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer:
J. Zahra

Telephone No. (41-22) 338.83.38

PATENT COOPERATION TREATY

BAKER BOTTS L.L.P.

01 FEB 16 AM 10:51

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To: HENRY TANG
BAKER BOTTS LLP
20 ROCKEFELLER PLAZA
NEW YORK, NY 10112-0228

PCT

TO

WRITTEN OPINION

(PCT Rule 66)

MCV

Date of Mailing *(day/month/year)*	12 FEB 2001
Applicant's or agent's file reference: 32312-PCT	**REPLY DUE** within **TWO** months from the above date of mailing

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US00/04505	22 FEBRUARY 2000	19 FEBRUARY 1999

International Patent Classification (IPC) or both national classification and IPC
IPC(7): G06F 17/27 and US Cl.: 707/500, 501, 530; 704/9, 10

Applicant
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

1. This written opinion is the first (first, etc.) drawn by this International Preliminary Examining Authority.

2. This opinion contains indications relating to the following items:

- I [X] Basis of the opinion
- II [ ] Priority
- III [ ] Non-establishment of opinion with regard to novelty, inventive step or industrial applicability
- IV [ ] Lack of unity of invention
- V [X] Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI [ ] Certain documents cited
- VII [ ] Certain defects in the international application
- VIII [ ] Certain observations on the international application

Docketed
For 2/12/2001 By

3. The applicant is hereby invited to reply to this opinion.

When? See the time limit indicated above. ~~The applicant may, before the expiration of that time limit, request this Authority to grant an extension., see Rule 66.2(d).~~

How? By submitting a written reply, accompanied, where appropriate, by amendments, according to Rule 66.3. For the form and the language of the amendments, see Rules 66.8 and 66.9.

Also For an additional opportunity to submit amendments, see Rule 66.4.
For the examiner's obligation to consider amendments and/or arguments, see Rule 66.4 *bis*.
For an informal communication with the examiner, see Rule 66.6.

If no reply is filed, the international preliminary examination report will be established on the basis of this opinion.

4. The final date by which the international preliminary examination report must be established according to Rule 69.2 is: 19 JUNE 2001.

Name and mailing address of the IPEA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	STEPHEN HONG James R. Matthews
Facsimile No. (703) 305-3230	Telephone No. (703) 305-3900

Form PCT/IPEA/408 (cover sheet) (July 1998)★

WRITTEN OPINION

International application No.
PCT/US00/04505

I. Basis of the opinion

1. With regard to the elements of the international application:*

[X] the international application as originally filed

[X] the description:
pages 1-15 , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of

[X] the claims:
pages 16-22 , as originally filed
pages NONE , as amended (together with any statement) under Article 19
pages NONE , filed with the demand
pages NONE , filed with the letter of

[X] the drawings:
pages 1-7 , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of

[X] the sequence listing part of the description: NONE , as originally filed
pages NONE , filed with the demand
pages NONE , filed with the letter of

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.
These elements were available or furnished to this Authority in the following language _____ which is:

[ ] the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).

[ ] the language of publication of the international application (under Rule 48.3(b)).

[ ] the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and or 55.3).

3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the written opinion was drawn on the basis of the sequence listing:

[ ] contained in the international application in printed form.

[ ] filed together with the international application in computer readable form.

[ ] furnished subsequently to this Authority in written form.

[ ] furnished subsequently to this Authority in computer readable form.

[ ] The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in international application as filed has been furnished.

[ ] The statement that the information recorded in computer readable form is identical to the writen sequence listing h been furnished.

4. [X] The amendments have resulted in the cancellation of:

[X] the description, pages NONE

[X] the claims, Nos. NONE

[X] the drawings, sheets/fig NONE

5. [ ] This opinion has been drawn as if (some of) the amendments had not been made, since they have been considered to beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

** Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referre in this opinion as "originally filed".*

WRITTEN OPINION

International application No.
PCT/US00/04505

V. Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. statement

Novelty (N)	Claims	1-37	YES
	Claims	NONE	NO
Inventive Step (IS)	Claims	NONE	YES
	Claims	1-37	NO
Industrial Applicability (IA)	Claims	1-37	YES
	Claims	NONE	NO

2. citations and explanations

Claims 1-37 lack an inventive step under PCT Article 33(3) as being obvious over Doi in view of Kupiec et al.

As per claims 1-37, Doi teaches the claimed feature of generating a summary by extracting sentences from the input document and parsing the extracted sentences into components (col.1, lines 52-60); sentence reduction processing which is performed to mark components which can be removed from the parsed sentences (FIG.4(B); col.3, lines 45-67); evaluating the importance of the context of the sentences and linguistic knowledge based processing (see the parts of speech analysis in FIG.3); combining sentences for identifying sentence combination operations and establishing rules for applying the sentence combination operations to merge at least two sentences and removing the unwanted portions of the sentences (col.5, line 20-35); and generating a summary of the document (see FIG.8; col.5, lines 35-42). However, Doi does not explicitly teach the use of the probablistic importance processing. Nevertheless, Kupiec et al. shows the probablistic processing for evaluating the importance in the summary generation system (col.1, lines 57 to col.2, line 17, "...the probability of observing a value of a particular feature in a sentence included in the summary and the probability of that feature taking each of its possible values..."). It would have been obvious to a person of ordinary skill in the art at the time of the invention to have incorporated Kupiec's probablistic processing into Doi, since Kupiec provided the motivation by pointing out that the probablistic model ensures more important parts of the sentences to be chosen for the summary.

Furthermore, although the prior art does not explicitly disclose the use of the "Hidden Markov Model" or "Viterbi algorithm" for the probability model, such were well known in the art, and thus, would have been obvious to a person of ordinary skill in the art at the time of the invention.

---------- NEW CITATIONS ----------

NONE

WRITTEN OPINION

International application No.
PCT/US00/04505

Supplemental Box
(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: Boxes I - VIII

TIME LIMIT:

The time limit set for response to a Written Opinion may not be extended. 37 CFR 1.484(d). Any response received after the expiration of the time limit set in the Written Opinion will not be considered in preparing the International Preliminary Examination Report.

32312-PCT 070050.1162

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

PCT RECEIVING OFFICE

Applicant	:	The Trustees of Columbia University in the City of New York
International Application No.	:	PCT/US00/04505
International Filing Date	:	22 February 2000
Title of Invention	:	CUT AND PASTE DOCUMENT SUMMARIZATION AND METHOD

EXPRESS MAIL LABEL NO. EF321686830US

RESPONSE TO FIRST WRITTEN OPINION

Commissioner for Patents and Trademarks
Box PCT
Washington, D.C. 20231

Attention: Stephen Hong
Authorized Officer
IPEA/US

Dear Sir:

Applicants respectfully respond the First Written Opinion which was mailed on February 12, 2001. In the First Written Opinion, the Authorized Officer indicated that, while novel and possessing industrial applicability, Claims 1-37 allegedly fail to present an inventive step over U.S. Patent 5,077,668 to Doi in view of U.S. Patent 5,778,397 to Kupiec et al. For the reasons set forth below, Applicants respectfully traverse this rejection and respectfully request issuance of a favorable written opinion in this case.

The present systems and methods which are described and claimed in the present application are directed to document summary generation using automated cut and paste operations. In general, the systems and methods operate by analyzing a document to identify one or more foci of the document and extract those sentences which are related to the document focus. The extracted sentences are reduced to remove language which is not critical to the resulting summary and can be combined such that multiple sentences of a document can be a single sentence in the resulting summary. The operations of sentence reduction and sentence combination as described and claimed in the present invention are not disclosed nor suggested by the prior art.

Referring to Claim 1, the Doi patent does not disclose extracting sentences related to a focus of a document, a grammatical parser, or a corpus of human generated summaries coupled to a sentence generation module. Rather than determining a document focus and than extracting sentences based on a relationship to that focus, Doi employs a table of "hint words" to identify sentences which may be important to a document. The "hint words" are generalized and not document specific. Thus, in Doi, if a sentence includes a "hint word" it is extracted, which can result in a large number of irrelevant sentences being extracted for a given summary.

In the present invention, the grammatical parser evaluates the extracted sentences to establish a grammatical representation of the words which make up the extracted sentence. In one case, a parse tree can be used to store the representation. (See Fig. 3). In contrast, Doi only identifies the parts of speech of the "hint words" which have been identified in an extracted sentence. Thus, Doi does not disclose employing a grammatical representation of the extracted sentences, such as a parse tree.

Another element of Claim 1 is the inclusion of a corpus of human generated summaries coupled to a sentence generation module. In Doi, a small set of fixed rules are identified for altering an extracted sentence by an Abstract Modification Unit. However,

there is no disclosure in Doi that the abstract modification unit is coupled to a corpus of human generated summaries or that the processing related to sentence modification is effected by an analysis of such a corpus. Accordingly, there is a substantial difference from the system of Claim 1 and that described in the Doi reference. In addition, the Kupiec reference does not provide any teaching to overcome the noted shortcomings of the Doi reference, with respect to Claim 1.

Claims 7-11 and 17-20 further define Claim 1 by specifying that the sentence generation module further comprises a sentence combination module for combining two extracted sentences in accordance with combination rules. In the Written Opinion, the Authorized Officer states that "combining sentences for identifying sentence combination operations and establishing rules for applying the sentence combination operations to merge at least two sentences..." is disclosed in Doi, Col. 5, lines 20-35. Applicants respectfully disagree. Throughout Doi, including the cited passage, operation is limited to a single sentence at any time. There is simply no disclosure regarding analyzing multiple extracted sentences and combining such sentences to form a new summary sentence. Each example of sentence modification recited in Col. 5, lines 20-35 discusses modifying a single extracted sentence and replacing the original extracted sentence with the new sentence. There is simply nothing to teach or suggest combining multiple sentences in the manner described and claimed in the present application.

The Doi patent does not teach or suggest the sentence reduction module of Claims 2-6 and Claims 13-16. The sentence modification disclosed in Doi is directed to a limited number of sentence transformation operations which take place on an extracted sentence based on the nature of the "hint words" in the extracted sentence. To the contrary, the present sentence reduction module evaluates the grammatical representation of the parsed extracted sentences and performs probabilistic importance processing (see, e.g., Claim 3), context importance processing (see, e.g., Claims 4 and 5) and relative component importance processing (see, e.g., Claim 6). Such processing involves rules based analysis based on a

combined lexicon and/or the corpus of human generated summaries. Referring to Claim 5, the context importance processing further includes establishing a plurality of lexical links among the sentence components and determining context importance based on such links. It is respectfully submitted that such elements of the claimed invention are not disclosed or suggested by the art of record.

The method of Claims 22-31 also recite the distinguishing features set forth above including sentence reduction and sentence combination of at least two sentences which can be merged. It is respectfully submitted that such claims also define an inventive step over the art of record.

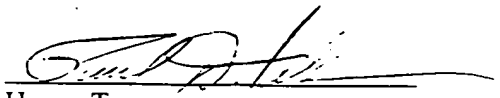
Claim 32 recites a method of identifying correspondence between phrases in a summary and phrases in the original document using a probability model. As set forth from page 13, line 18 to page 15, line 24, this method is particularly applicable to performing an analysis of a corpus of summaries for establishing new rules or partitioning the corpus into sub corpora. The art of record only discloses summary generation and does not disclose or suggest analyzing a summary to determine its relationship to an original document. Accordingly, claims 32-34 define an inventive step over the art of record.

Claims 35-37 are directed to a corpus for a summarization system which includes a plurality of documents, a plurality of human generated summaries of the documents, a sentence combination subcorpus and a sentence reduction subcorpus. As set forth above, the art of record does not disclose the use of a corpus of human generated summaries and is silent as to sentence combination as performed by the present invention. Accordingly, the corpus which is defined by Claims 35-37 represents an inventive step over the art of record.

In view of the foregoing remarks, reconsideration of the First Written Opinion and issuance of a favorable Second Written Opinion with respect to Claims 1-37 is respectfully solicited.

BAKER BOTTS L.L.P.

Dated: April 11, 2001



Henry Tang
Reg. No. 29,705

Paul D. Ackerman
Reg. No. 39,891

Attorneys for Applicant
(212) 705-5000

Enclosures

09/913946

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 32312-PCT	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/US00/04505	International filing date *(day/month/year)* 22 FEBRUARY 2000	Priority date *(day/month/year)* 19 FEBRUARY 1999
International Patent Classification (IPC) or national classification and IPC IPC(7): G06F 17/27 and US Cl.: 707/500, 501, 530; 704/9, 10		
Applicant THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 4 sheets.

   ☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority. (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of — sheets.

3. This report contains indications relating to the following items:

CORRECTED VERSION

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of report with regard to novelty, inventive step or industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
05 SEPTEMBER 2000	07 APRIL 2001
Name and mailing address of the IPEA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHEN HONG Telephone No. (703) 305-3900

Form PCT/IPEA/409 (cover sheet) (July 1998)★

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US00/04505

I. Basis of the report

1. With regard to the elements of the international application:*

[X] the international application as originally filed

[X] the description:
pages 1-15, as originally filed
pages NONE, filed with the demand
pages NONE, filed with the letter of _____

[X] the claims:
pages 16-22, as originally filed
pages NONE, as amended (together with any statement) under Article 19
pages NONE, filed with the demand
pages NONE, filed with the letter of _____

[X] the drawings:
pages 1-7, as originally filed
pages NONE, filed with the demand
pages NONE, filed with the letter of _____

[X] the sequence listing part of the description:
pages NONE, as originally filed
pages NONE, filed with the demand
pages NONE, filed with the letter of _____

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.
These elements were available or furnished to this Authority in the following language _____ which is:

[ ] the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).

[ ] the language of publication of the international application (under Rule 48.3(b)).

[ ] the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

[ ] contained in the international application in printed form.

[ ] filed together with the international application in computer readable form.

[ ] furnished subsequently to this Authority in written form.

[ ] furnished subsequently to this Authority in computer readable form.

[ ] The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

[ ] The statement that the information recorded in computer readable form is identical to the writen sequence listing has been furnished.

4. [X] The amendments have resulted in the cancellation of:

[X] the description, pages NONE

[X] the claims, Nos. NONE

[X] the drawings, sheets/~~fig~~ NONE

5. [ ] This report has been drawn as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

** Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to*

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US00/04505

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. statement

Novelty (N)	Claims	1-37	YES
	Claims	NONE	NO
Inventive Step (IS)	Claims	NONE	YES
	Claims	1-37	NO
Industrial Applicability (IA)	Claims	1-37	YES
	Claims	NONE	NO

2. citations and explanations (Rule 70.7)

Examiner's Response to the Applicant's arguments filed on 11 April 2001.

On page 4 of the argument, Applicant asserts that Doi patent does not disclose "extracting sentences related to a focus of a document, a grammatical parser or a corpus of human generated summaries coupled to a sentence generation module." Examiner disagrees.

Referring to claim 1, nothing limits Doi's "hint word" as the focus of a document from which the sentences are extract. Furthermore, Doi discloses the grammatical parser, since the summaries are in fact generated by combining multiple sentences from the parsed content.

Lastly, Doi teaches the human generated a collection of "corpus" of human summaries, since Doi teaches the operator edited summary sentences coupled to the generation module (col.1, line 52).

Claims 1-37 lack an inventive step under PCT Article 33(3) as being obvious over Doi in view of Kupiec et al.

As per claims 1-37, Doi teaches the claimed feature of generating a summary by extracting sentences from the input document and parsing the extracted sentences into components (col.1, lines 52-60); sentence reduction processing which is performed to mark components which can be removed from the parsed sentences (FIG.4(B); col.3, lines 45-67); evaluating the importance of the context of the sentences and linguistic knowledge based processing (see the parts of speech analysis in FIG.3); combining sentences for identifying sentence combination operations and establishing rules for applying the sentence combination operations to merge at least two sentences and removing the unwanted portions of the sentences (col.5, line 20-35); and generating a summary of the document (see FIG.8; col.5, lines 35-42). However, Doi does not explicitly teach the use of the probablistic importance processing. Nevertheless, Kupiec et al. shows the probablistic processing for evaluating the

(Continued on Supplemental Sheet.)

Supplemental Box
(To be used when the space in any of the preceding boxes is not sufficient)

V. 2. REASONED STATEMENTS - CITATIONS AND EXPLANATIONS (Continued):
importance in the summary generation system (col.1, lines 57 to col.2, line 17, "...the probability of observing a value of a particular feature in a sentence included in the summary and the probability of that feature taking each of its possible values..."). It would have been obvious to a person of ordinary skill in the art at the time of the invention to have incorporated Kupiec's probablistic processing into Doi, since Kupiec provided the motivation by pointing out that the probablistic model ensures more important parts of the sentences to be chosen for the summary.

Furthermore, although the prior art does not explicitly disclose the use of the "Hidden Markov Model" or "Viterbi algorithm" for the probability model, such were well known in the art, and thus, would have been obvious to a person of ordinary skill in the art at the time of the invention.

---------- NEW CITATIONS ----------

NONE

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 January 2001 (25.01.2001)

PCT

(10) International Publication Number
WO 01/06408 A1

(51) International Patent Classification[7]: G06F 17/27

(21) International Application Number: PCT/US00/04505

(22) International Filing Date: 22 February 2000 (22.02.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/120,657 19 February 1999 (19.02.1999) US

(71) Applicant *(for all designated States except US)*: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 116th Street and Broadway, New York, NY 10027 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MCKEOWN, Kathleen, R.** [US/US]; 20 Prospect Road, Wayne, NJ 07470 (US). **JING, Hongyan** [US/US]; 521 West 112th Street, Apt. 73C, New York, NY 10025 (US).

(74) Agents: TANG, Henry et al.; Baker & Botts LLP, 30 Rockefeller Plaza, New York, NY 10112-0228 (US).

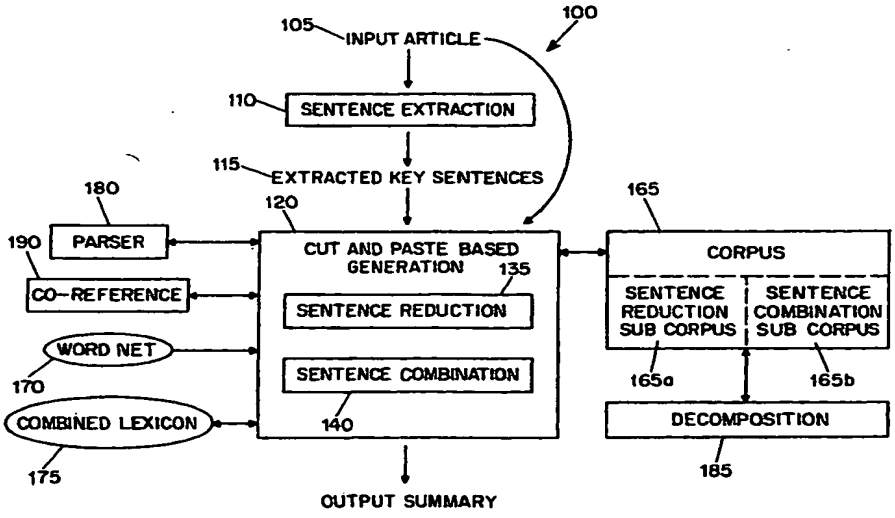
(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CUT AND PASTE DOCUMENT SUMMARIZATION SYSTEM AND METHOD



(57) Abstract: A summary of an input document is generated by extracting at least one sentence from the document and parsing the extracted sentences into components, such as in a parse tree (110). Sentence reduction processing is performed to mark components which can be removed from the parse trees (135). Sentence reduction can include context importance processing, probabilistic processing, and linguistic knowledge based processing, probabilistic processing includes identifying sentence combination operations and establishing rules for applying the sentence combination operations to mark the parse trees to merge at least two sentences (140). Sentence combination processing also provides a paste operation to operate on the marked components to effect the indicated removal and combination of sentence components, thereby generating summary sentences for the input document.

WO 01/06408 A1

CUT AND PASTE DOCUMENT SUMMARIZATION SYSTEM AND METHOD

Statement of Government Rights

The United States Government may have certain rights to the invention set forth herein pursuant to a grant by the National Science Foundation, Contract No. IRI-96-198124

Statement of Related Applications

This application claims the benefit of United States provisional patent application, Serial No. 60/120,657, entitled "Summary Generation Through Intelligent Cutting and Pasting of the Input Document" which was filed on February 19, 1999.

Field of the Invention

The present invention relates generally to information summarization and more particularly relates to systems and methods for generating a summary of a document using automated cutting and pasting of the input document.

Background of the Invention

The amount of information available today drastically exceeds that of any other time in history. With the continuing expansion of the Internet, this trend will likely continue well into the future. Often, people conducting research of a topic are faced with information overload as the number of potentially relevant documents exceeds the researcher's ability to individually review each document. To address this problem, information summaries are often relied on by researchers to quickly evaluate a document to determine if it is truly relevant to the problem at hand.

Given the vast collection of documents available, there is interest in developing and improving the systems and methods used to summarize information content. For individual documents, domain-dependent template based systems and

domain-independent sentence extraction methods are known. Such known systems can provide a reasonable summary of a single document when the domain is known.

Many presently available summarizers extract sentences from the original documents to produce summaries. However, since the sentences are generally extracted without supporting context information, the resulting summaries can be incoherent, and in some cases, can convey misleading information.

Therefore, there remains a need for systems and methods which can generate a more readable and concise summary of a document.

Summary of the Invention

It is an object of the present invention to provide a system and method for generating a summary of a document.

It is another object of the present invention to provide a summarization system which extracts sentences from an input document and then transforms the extracted sentences such that a concise, coherent and accurate summary results.

It is a further object of the present invention to provide a system and method for generating a summary of a set document which use automated cutting and pasting of the input document.

A present method for generating a summary of an input document includes extracting at least one sentence from the document. The extracted sentences are parsed into components, preferably in a parse tree representation. Sentence reduction is performed to mark components which can be removed from the extracted sentences. Sentence combination is performed to mark components of two or more sentences which can be merged. Sentence combination also includes a paste operation to operate on the marked components to effect the indicated removal and combination of sentence components.

A preferred sentence reduction operation includes measuring the contextual importance of the components; measuring the probabilistic importance of the components based on a given corpus; measuring the importance of the components based on linguistic knowledge; synthesizing the contextual, probabilistic and knowledge based importance measures into a relative importance score for each

component; and marking for removal those components with an importance score below a threshold value.

The contextual importance can be measured by establishing a plurality of lexical links of at least one type among the components in a local context in the document and computing a context importance score according to the type, number and direction of lexical links associated with each component. The types of lexical links can include repetition, inflectional variants, derivational variants, synonyms, hypernyms, antonyms, part-of, entailment, and causative links.

In a preferred method, the sentence combination operation includes identifying sentence combination operations from a sentence combination subcorpus and developing rules regarding the application of the sentence combination operations. The combination rules are then applied to the extracted sentences after sentence reduction to identify and merge suitable sentences from the original article. The sentence combination operations can be selected from the group including add descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

A present system for generating a summary of an input document includes an extraction module which receives the input document and extracts at least one sentence related to a focus of the document. A summary sentence generation module is provided, which generally includes a sentence reduction module and a sentence combination module. The system includes a grammatical parser operatively coupled to the generation module for parsing the extracted sentences into components in a grammatical representation. A combined lexicon and a corpus of human generated summaries are operatively coupled to the generation module for use by the operational modules during summary generation.

The corpus can further include a sentence generation subcorpus and a sentence reduction subcorpus. The subcorpora can be generated manually or through the use of a decomposition module.

Preferably, the sentence reduction module is cooperatively engaged with the combined lexicon and performs context importance processing on the components of the grammatical representation. Context importance processing can

include establishing a plurality of lexical links of at least one type for the components and generating a context importance score based on the type and number of links associated with the components. The number and type of lexical links can vary, however a preferred set of lexical link types includes repetition, inflectional variants, derivational variants, synonyms, hypernyms, antonyms, part-of, entailment, and causative links.

Preferably, the sentence reduction module further computes the relative importance of the components based on linguistic knowledge stored in the combined lexicon. The sentence reduction module can also be cooperatively engaged with the corpus and perform probabilistic importance processing on the components of the grammatical representation in accordance with the particular corpus used.

The sentence combination module can be used to identify sentence combination operations from a sentence combination subcorpus and develop rules regarding the application of the sentence combination operations. The combination module applies the combination rules to the extracted sentences after sentence reduction to identify and merge suitable sentences from the original article.

A decomposition module in accordance with the present application can be used to evaluate human generated summaries and map corresponding portions of the summaries to the original documents. The decomposition module indexes words in the summary and the original document. A Hidden Markov Model is then built based on heuristic rules to determine the probability of phrases in the summary sentence matching a given phrase in the original document. A Viterbi algorithm can then be employed to determine the best solution for the Hidden Markov Model and generate a mapping between summary phrases and the original document. This mapping can be used to generate, among other things, a sentence reduction subcorpus and a sentence combination subcorpus. Such a decomposition module can be operatively coupled to the corpus in the summary generation system described above.

Brief Description of the Drawing

Further objects, features and advantages of the invention will become apparent from the following detailed description taken in conjunction with the accompanying figures showing illustrative embodiments of the invention, in which

Figure 1 is a block diagram of the system architecture of the present document summarization system;

Figure 2 is a flow chart illustrating an exemplary embodiment of a sentence reduction operation in accordance with the summarization system of Figure 1;

Figure 3 is a pictorial diagram of an exemplary parse tree sentence representation;

Figure 4 is a flow chart illustrating an exemplary embodiment of a sentence combination operation in accordance with the present summarization system of Figure 1;

Figure 5 is a table illustrating exemplary sentence combination operations for the sentence combination operation of Figure 4;

Figure 6 is a table illustrating exemplary sentence combination rules for applying the sentence combination operations of Figure 5;

Figure 7 is a flow diagram illustrating the operation of the corpus decomposition module of Figure 1; and

Figure 8 is a pictorial diagram of a Hidden Markov Model for use in a corpus decomposition module.

Throughout the figures, the same reference numerals and characters, unless otherwise stated, are used to denote like features, elements, components or portions of the illustrated embodiments. Moreover, while the subject invention will now be described in detail with reference to the figures, it is done so in connection with the illustrative embodiments. It is intended that changes and modifications can be made to the described embodiments without departing from the true scope and spirit of the subject invention as defined by the appended claims.

Detailed Description of Preferred Embodiments

The present summarization systems and methods generate a generic, domain-independent single-document summary of a received input document. Figure 1 is a block diagram illustrating the system architecture of an exemplary embodiment of the present summarization system. Such a system can be implemented on various computer hardware, software and operating system platforms. The particular system components selected are not critical to the practice of the present invention. For example, the present system of Figure 1, can be implemented on a personal computer system, such as an IBM compatible system.

Referring to Figure 1, an input document 105 in computer readable form is applied to an extraction module 110 which determines the focus of the document 105 and extracts sentences from the document accordingly. A number of extraction techniques can be used in the extraction module 110. In a preferred embodiment, the extraction module 110 links words in a sentence to other words in the input document 105 through repetitions, morphological relations and lexical relations. An importance score can then be computed for each word in the article 105 based on the number, type and direction (forward, backward) of the lexical links associated with the word. A sentence score can be determined by adding the importance score for each of the words in the sentence and normalizing the sum based on the number of words in the sentence. The sentences can then extracted based on the highest relative sentence scores.

The extraction module 110 provides the extracted sentences 115 to a generation module 120. The generation module 120 also receives the original document 105 as an input. The generation module 120 further includes a sentence reduction module 135 and a sentence combination module 140. The sentence reduction module 135 provides a marked up parse tree as input data for the sentence combination module 140, which generates and outputs the summary sentences.

The generation module 120 is operatively coupled to a corpus of human-written summaries 165, a lexical database 170, and a combined reusable lexicon 175.

The corpus 165 generally includes a broad collection of human-generated summaries as well as the corresponding original documents. The corpus 165 can also include a sentence reduction subcorpus 165a and a sentence combination subcorpus 165b which can be generated manually or through a decomposition module. The sentence reduction subcorpus 165a includes entries of sentence pairs linking an original sentence to a human reduced sentence. The sentence combination subcorpus 165b includes mappings from human combined sentences to two or more original sentences.

A suitable exemplary corpus 165 was generated using *Communications-related Headlines*, a free daily online news service provided by the Benton Foundation (http://www.benton.org). The articles from this particular service are communication related, but the topics involved are very broad, including law, company mergers, new technologies, labor issues and so on. Of course, other sources of document summaries can also be used to generate a suitable corpus. To insure that the resulting corpus is somewhat generic, the articles from the selected source should not possess a particular writing style. Thus, preferred sources feature articles from multiple sources or articles from various sections of one or more source. A suitable corpus 165 was generated in four major steps. First, human-written, single document summaries are received from the source. Second, the original documents are retrieved and correlated to the respective summary. The retrieved documents are then "cleaned" by removing irrelevant material such as indexes and advertisements. Finally, the quality of the correspondence between the summary and the original document is verified. The cleaning and verification processes are generally performed manually. The sentence reduction subcorpus 165a and sentence combination subcorpus 165b entries were generated by the decomposition module 185, the operation of which is explained below.

The lexical database 170 can take the form of the WordNet database, which is described in the article "WordNet: A lexical Database for English", by G.A. Miller, Communications of the ACM, Vol. 38, No. 11. pp. 39-41, November 1995. A suitable embodiment of the combined lexicon 175 can be constructed by combining multiple, large-scale resources such as WordNet, the English Verb Classes and

Alternations (EVCA) database, the COMLEX syntax dictionary and the Brown Corpus tagged with WordNet senses. The combined lexicon 175 can be formed by encoding the EVCA database with COMLEX compatible syntax and merging the EVCA into the COMLEX database. This results in each verb in the combined lexicon 175 being marked with a list of subcategorizations and alternate syntactic patterns. Preferably, WordNet is added to the EVCA/COMLEX combination to refine the syntactic information and provide additional lexical information to the lexicon 175.

The generation module 120 is also cooperatively coupled to natural language processing (NLP) tools such as a syntactic parser 180 and a co-reference resolving module 190 which can include anaphora resolution. These tools can be software modules which are called by the generation module 120. A suitable syntactic parser 180 is the English Slot Grammar (ESG) parser available from International Business Machines, Inc. A suitable co-reference resolving module 190 is the Deep Read system, available from Mitre, Inc.

Figure 2 is a flow diagram further illustrating the operation of the sentence reduction module 135. The reduction module 135 receives extracted sentences 115 as input (step 205). The reduction module invokes the parser 180 to grammatically parse the extracted sentences 115 and generate a parse tree representation of the sentences (step 210). In step 215 contextual importance is determined by detecting lexical links among words in a local context and then computing an importance score based on the number, type and direction of lexical links detected. The context processing step 215 generates an importance score for each node in the parse tree indicating the relative importance of the nodes to the focus of the input document 105.

The number, type and direction (forward, backward) of lexical links used in the practice of the present invention may vary. An empirical study has demonstrated that the following nine lexical relation types provide a meaningful representation of contextual importance: (1) repetition, (2) inflectional variants, (3) derivational variants, (4) synonyms, (5) hypernyms, (6) antonyms, (7) part-of, (8) entailment (for example: *kill* → *die*), and (9) causative (for example: *eat* → *chew*).

Inflectional variants (2) and derivational variants can be derived from the CELEX database content, available from the Centre for Lexical Information, Max Planck Institute for Psycholinquistics, Nijmegen, which can be in the combined lexicon 175. The other lexical relations can be extracted using the separate lexical database 170, such as WordNet. To frame the local context of a word, a number of sentences before and after the current sentence location are evaluated for the presence of lexical links. The number of sentences selected for this operation involves balancing the level of contextual depth to the amount of processing overhead. Using the five sentences before and the five sentences after the current sentence has been found to provide reasonable local context without incurring excessive processing overhead.

After the lexical links have been identified (step 215 a), an importance score for each word in the extracted sentences can be calculated (step 215 b). Lexical links from the current sentence to subsequent sentences are referred to as forward links and those from the current sentence to preceding sentences are referred to as backward links. The importance score, referred to as the context weight, can be computed as follows:

$$1)\quad ForwardWeight(w) = \sum_{i=1}^{9} (WixLi(w))$$

$$2)\quad BackwardWeight(w) = \sum_{i=1}^{9} (WixBnumi(w))$$

$$3)\quad TotalWeight(w) = ForwardWeight(w) + BackwardWeight(w)$$

$$4)\quad Ratio(w) = \frac{\max(ForwardWeight(w), Backward(weight(w))}{Totalweight(w))}$$

$$5)\quad ContextWeight = Ratio(w) x TotalWeight(w)$$

where *ForwardWeight(w)* computes the weight of forward links, *BackwardWeight(w)* computes the weight of backward links, *TotalWeight*(w) represents the sum of all links and *Ratio*(w) computes a weight for the location of the word. To compute the weight of various lexical links, each type of link is assigned a weighted value according to its relative importance. For example, the nine lexical relations set forth

above were presented in descending order of importance and accordingly can be assigned linearly decreasing weights such as (1,0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2). The value of *Ratio(w)* represents the value assigned based on the location of the word in the original document. For example, when a sentence introduces a topic or ends a topic, it is considered more important and the components of those sentences will be assigned a relatively higher location value.

The use of various types of lexical relations improves the relatedness of a word to the main topic. Although simple relations like repetition and synonymy can be used to determine a measure of contextual importance, these surface relations are generally unable to detect more subtle connections between words.

Following context processing (step 215) the reduction module 135 can perform interdependency processing using a probability analysis based on the corpus 165 of human-written reduction based sentences. Such an analysis can indicate the degree of correlation between components in a sentence, such as the relationship between a verb and its subclause.

The probability computation can be performed based on parse trees using probabilities to indicate the degree of correlation between a parent node and its child nodes in the parse tree. Figure 3 illustrates an exemplary fragment of a parse tree used to explain the operation of the probability computation. In Fig. 3, The main verb "give" is the parent nodes 300, and it has four children nodes: subclause conjunct 305, subject 310, indirect object 315 and object 320, respectively. The parse tree can also include further levels below the children nodes, such as nodes ndet 325 and adjp 330 below child node obj 320 and nodes lconj 335 and rconj 340 below node adjp 330, respectively.

To measure the interdependency between the verb *give* and its subclause 305, the probability that the subclause is removed when the verb is *give*, can be represented by *PROB("when_clause is removed"|verb = give)* . This conditional probability is transformed using Bayes's rule to:

$$PROB(\text{"when\_clause is removed"} \mid \text{"v = give"}) = \frac{PROB(\text{"v = give"} \mid \text{"when\_clause removed"}) * PROB(\text{"when\_clause is removed"})}{PROB(\text{verb = give})}$$

In a similar fashion, the probabilities that a clause will be reduced or remain unchanged can be calculated in a similar manner.

The probability associated with the other child nodes from the current root node is calculated in a similar manner. After the probabilities for each of the first level child nodes is calculated, each of the child nodes in the current level of the tree is then treated as a parent node and the process is repeated through each descending level of the parse tree until every parent-child node pair has been considered. The probabilities for the corpus 165 can be calculated and stored in a look-up table which is used when a reduction module 135 is run.

The context processing of step 215 and probability processing of step 220 provide a relative ranking of sentence components. However, this ranking does not necessarily provide a measure of which components be included to provide a grammatically correct summary sentence. Thus, preferably, after the probability analysis of step 220, reduction processing based on linguistic knowledge is performed (step 225). In this operation, the reduction module 135 works in cooperation with the combined lexicon 175.

The linguistic knowledge processing step 225 operates with the combined lexicon 175 to evaluate the parse tree for each extracted sentence 115 and determine which children nodes are essential to maintain the grammatical correctness of the component represented by the parent node. Linguistic judgments are identified in the parse tree by assigning a binary tag to each node in the parse tree. The value of a tag is either *essential* or *reducible*, indicating whether or not a node is indispensable to its parent node. For example, referring to Figure 3, the lexicon 175 will indicate that the verb *give* needs a subject and two objects. Thus the child nodes *subj 310, iobj 315* and *obj* 320 can be marked as essential. In this case, the child node subclause 305 is then rendered non-essential and will be marked as reducible. The lexicon 175 can also include collocations, such as *consist of* or *replace .... with .....*, which prevents removal of indispensable components.

Once the linguistic knowledge processing is applied in step 225, a reduction operation (step 230) can take place. The reduction operation process can be viewed as a series of decision making steps along the edges of a parse tree. Beginning

with the root node of the parse tree, the immediate child nodes are evaluated to determine which child nodes can be removed. A child node can be removed if three conditions are satisfied. The first condition is that the component is not a local focus. To determine whether a component is a local focus, the ratio of the context importance score (step 215b) of the child node to that of the root node is calculated. The child node is then considered unimportant if the calculated ratio is smaller than a threshold value. The second condition is that the corpus probability value (step 220) indicating that the special syntactic component of the root is removed is higher than a threshold. The final condition is that the linguistic analysis in step 225 indicates that the child node as reducible.

When the conditions to remove a child node are satisfied, the child node is tagged as "removable" and processing on that branch of the tree terminates. For the child nodes which are retained, the lower levels of the parse tree are evaluated by repeating this process in a similar manner through the tree. The reduction operation step 230 is complete when there are no more nodes to consider. This also concludes processing of the sentence reduction module and results in the parse trees being marked with those components which can be removed or altered by the subsequent paste module 150 operation.

Following processing by the sentence reduction module 135, processing by the sentence combination module 140 is performed. The operation of the sentence combination module 140 is further illustrated in the flow chart of Figure 4.

Using the sentence combination subcorpus165b, the sentence combination module evaluates the extracted sentence to identify applicable sentence combination operations (step 410). Figure 5 is a table illustrating combination operations such as: add descriptions 510, aggregations 515, substitute incoherent phrases 520, substitute phrases with more general or more specific information 525 and mixed operations 530.

From the sentence combination subcorpus 165b, sentence combination rules are also established to determine whether and how the sentence combination operations of step 410 will take place (step 415). The result is a set of sentence

combination rules 420, such as those set forth in Figure 6. The rules illustrated in Figure 6 are exemplary and non-exhaustive. These sentence combination rules 420 were determined empirically by manual inspection of the sentence combination subcorpus 165b. Using the input article 105 and the extracted sentences reduced by the sentence reduction module 135 the sentence combination module 140 in cooperation with the co-reference resolution module 190 applies the sentence combination rules 420 (step 425). The result of step 425 is that the parse trees of the sentences being combined are appropriately tagged to effect the sentence combination. The combination operation is then realized in step 430 using a tree adjoining grammar (TAG) formalism, as described by A. Joshi, "Introduction to Tree-Adjoining Grammars," in Mathematics of Language, John Benjamins, Amsterdam, 1987. In this way, the sentence combination module 140 performs a paste operation on the marked parse trees and generates a summary sentence.

The document summary is generated by combining the summary sentences. The most straight forward combination is to maintain the order of sentences as they were extracted, however, other sequencing arrangements can also be employed.

As noted above in connection with Figure 1, the corpus decomposition module 185 operates on the corpus 165 to generate the sentence reduction subcorpus 165a and the sentence combination subcorpus 165b. The decomposition module 185 generally operates to evaluate the human written summaries in the corpus 165, compare the summary sentences to the original document, determine if a summary sentence was generated by a cut and past operation and identify where the components of the summary sentences were taken from in the original documents. The operation of the decomposition module 185 is illustrated in the flow diagram of Figure 7.

Referring to Figure 7, the decomposition module 185 uses the human-generated summary and original document as inputs to an indexing operation (step 705). During indexing, each word in the original document is indexed according to its positions in the original document. A convenient way of referencing these occurrences is by sentence number and word number in the original document.

To evaluate the index of words, a set of heuristic rules is developed by manual inspection of the corpus 165. Such inspection reveals that human-generated summaries often include one or more of six operations: sentence reduction, sentence combination, syntactic transformation, lexical paraphrasing, generalization/specification, and content reordering. The heuristic rules can be represented using a bigram probability $PROB\,(W_2 = (S_2, W_2) | W_1 = (S_1, W_1))$ (abbreviated as $PROB(W_2|W_1)$ in the following discussion). The probability values can be assigned in the following manner:

- $IF\ ((S_1 = S_2) and (W_1 = W_2 - 1))$ (i.e., the words are in two adjacent positions in the document), $THEN\ PROB(W_2|W_1)$ is assigned the maximal value, P1.(Rule: Two adjacent words in the summary are most likely to come from two adjacent words in the document.)

- $IF\ ((S_1 = S_2) and (W_1 < W_2 - 1))$, $THEN\ PROB(W_2|W_1)$ is assigned the second highest value, P2. (Rule: Adjacent words in the summary are highly likely to come from the same sentence in the document, retaining their relative precedent relation, as in sentence reduction. This rule can be further refined by adding restrictions on distance between words.)

- $IF\ ((S_1 = S_2) and (W_1 > W_2))$, $THEN\ PROB(W_2|W_1)$ is assigned the third highest value, P3. (Rule: Adjacent words in the summary are likely to come from the same sentence in the document but reverse their relative orders, such as in the case of sentence reduction with syntactic transformations.)

- $IF\ (S_2 - CONST < S_1 < S_2)$, $THEN\ PROB(W_2|W_1)$ is assigned the fourth highest value, P4. (Rule: Adjacent words in the summary can come from nearby sentences in the document and retain their relative order, such as in sentence combination. CONST is a small constant such as 3 or 5.)

- $IF\ (S_2 < S_1 < S2 + CONST)$, $THEN\ PROB(W_2|W_1)$ is assigned the fifth highest value, P5. (Rule: Adjacent words in the summary can come from nearby sentences in the document but reverse their relative orders.)

- $IF\ (|S_2 - S_1| \geq CONST)\ THEN\ PROB(W_2|W_1)$ is assigned a small value, P6. (Rule: Adjacent words in the summary are not very likely to come from sentences far apart.)

Based on the above heuristic principles, a Hidden Markov Model can be generated, such as is illustrated in Figure 8 (step 710). The nodes in the Hidden Markov Model represent possible positions in the document, and the edges output the probability of going from one node to another. This Hidden Markov Model is used in finding the most likely position sequence in a subsequent processing operation. Assigning values to Pl-P6 is performed empirically. For example, the maximal value can be assigned 1 and others are assigned evenly decreasing values 0.9, 0.8 and so on. The order of the above rules is based on the empirical observations on a particular set of summaries. These values, however, can be adjusted or even trained for different corpora.

A Viterbi algorithm can be used to evaluate the Hidden Markov Model and find the most likely sequence of words incrementally (step 715). The Viterbi algorithm first finds the most likely sequence for ($Word_1Word_2$), for each possible position of $Word_2$. This information is then used to compute the most likely sequence for ($Word_1Word_2Word_3$), for each possible position of $Word_3$. The process repeats until all the words in the sequence have been considered.

After evaluation by the Viterbi algorithm, post-editing operations can be used to cancel mismatches that occur in the corpus analysis. The result is that summary sentences are matched to the corresponding phrases in the document. Once the summary sentences are so matched, it is a simple endeavor to sort the various matchings to one of the sentence reduction subcorpus 165a and sentence combination subcorpus 165b. In addition, the decomposition module 185 can be used as a stand alone tool, apart from the rest of the present summary generation system, to perform various summary analysis operations.

Although the present invention has been described in connection with specific exemplary embodiments, it should be understood that various changes, substitutions and alterations can be made to the disclosed embodiments without departing from the spirit and scope of the invention as set forth in the appended claims.

CLAIMS

1. A system for generating a summary of an input document comprising:

an extraction module, the extraction module receiving the input document and extracting at least one sentence related to a focus of the document;

a summary sentence generation module operatively coupled to the extraction module;

a grammatical parser operatively coupled to the generation module for parsing the extracted sentences into components in a grammatical representation;

a combined lexicon operatively coupled to the generation module; and

a corpus of human generated summaries operatively coupled to the generation module.

2. The system for generating a summary of an input document of claim 1, wherein the generation module further comprises a sentence reduction module.

3. The system for generating a summary of an input document of claim 2, wherein the sentence reduction module is cooperatively engaged with the corpus and performs probabilistic importance processing on the components of the grammatical representation in accordance with the corpus.

4. The system for generating a summary of an input document of claim 3, wherein the sentence reduction module is cooperatively engaged with the combined lexicon and performs context importance processing on the components of the grammatical representation.

5. The system for generating a summary of an input document of claim 4, wherein the context importance processing includes establishing a plurality of lexical links of a least one type for the components and generating a context importance score based on the type and number of links associated with the components.

6. The system for generating a summary of an input document of claim 5, wherein the sentence reduction module further computes the relative importance of the components based on linguistic knowledge stored in the combined lexicon.

7. The system for generating a summary of an input document of claim 1, wherein the generation module further comprises a sentence combination module.

8. The system for generating a summary of claim 7, wherein the sentence combination module is operatively coupled to the corpus and wherein the sentence combination module:

identifies at least one sentence combination operation;

establishes at least one rule for applying the sentence combination operation; and

applies the at least one rule to combine at least two extracted sentences.

9. The system for generating a summary of claim 8, wherein the at least one sentence combination operation is selected from the group consisting of add descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

10. The system for generating a summary of claim 9, wherein the at least one rule to combine extracted sentences includes replacing a partial name phrase with a full name phrase.

11. The method of generating a summary of claim 10, wherein the at least one rule to combine extracted sentences includes determining if two sentences having a common subject are proximate and whether at least one sentence is marked for reduction then removing the subject of the second sentence and combining with the first sentence using the connective "and."

12. The system for generating a summary of an input document of claim 1, wherein the generation module further comprises a sentence reduction module and a sentence combination module.

13. The system for generating a summary of an input document of claim 12, wherein the sentence reduction module is cooperatively engaged with the combined lexicon and performs context importance processing on the components of the grammatical representation.

14. The system for generating a summary of an input document of claim 13, wherein the context importance processing includes establishing a plurality of lexical links of a least one type for the components and generating a context importance score based on the type and number of links associated with the components.

15. The system for generating a summary of an input document of claim 14, wherein the sentence reduction module further computes the relative importance of the components based on linguistic knowledge stored in the combined lexicon.

16. The system for generating a summary of an input document of claim 15, wherein the sentence reduction module is cooperatively engaged with the corpus and performs probabilistic importance processing on the components of the grammatical representation in accordance with the corpus.

17. The system for generating a summary of an input document of claim 12, wherein the sentence combination module is operatively coupled to the corpus and wherein the sentence combination module:

identifies at least one sentence combination operation;

establishes at least one rule for applying the sentence combination operation; and

applies the at least one rule to combine at least two extracted sentences.

18. The system for generating a summary of claim 17, wherein the at least one sentence combination operation is selected from the group consisting of add descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

19. The system for generating a summary of claim 18, wherein the at least one rule to combine extracted sentences includes replacing a partial name phrase with a full name phrase.

20. The method of generating a summary of claim 19, wherein the at least one rule to combine extracted sentences includes determining if two sentences having a common subject are proximate and whether at least one sentence is marked for reduction then removing the subject of the second sentence and combining with the first sentence using the connective "and."

21. The system for generating a summary of an input document of claim 1, further comprising a decomposition module operatively coupled to the corpus, the decomposition module analyzing the corpus and generating a sentence reduction subcorpus and a sentence combination subcorpus.

22. A method of generating a summary of an input document comprising:

extracting at least one sentence from the document;

parsing the at least one sentence into components;

performing a sentence reduction operation to mark components which can be removed from the sentence;

performing a sentence combination operation to mark components of at least two sentences which can be merged; and

operating on the marked components to effect the indicated removal and combination of sentence components.

23. The method of generating a summary of claim 22, wherein the sentence reduction operation comprises:

measuring the contextual importance of the components;

measuring the probabilistic importance of the components based on a given corpus;

measuring the importance of the components based on linguistic knowledge;

synthesizing the contextual, probabilistic and knowledge based importance measures into a relative importance score for each component; and

marking those components having an importance score below a threshold value for removal.

24. The method of generating a summary of claim 23, wherein the contextual importance is measured by:

identifying a plurality of lexical links of at least one type among the components in a local context in the document; and

computing a content importance score according to the type and number of lexical links associated with each component.

25. The method of generating a summary of claim 24, wherein the at least one type of lexical links are selected from the group consisting of repetition, inflectional variants, derivational variants, synonyms, hypernyms, antonyms, part-of, entailment, and causative links.

26. The method of generating a summary of claim 23, wherein the probabilistic importance score is determined based on a corpus of human-written summaries.

27. The method of generating a summary of claim 23, wherein the linguistic knowledge operation includes the use of a combined lexicon.

28. The method of generating a summary of claim 22, wherein the sentence combination operation further comprises:

identifying at least one sentence combination operation;

establishing at least one rule for applying the sentence combination operation; and

applying the at least one rule to combine at least two extracted sentences.

29. The method of generating a summary of claim 28, wherein the at least one sentence combination operation is selected from the group consisting of add descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

30. The method of generating a summary of claim 28, wherein the at least one rule to combine extracted sentences includes replacing a partial name phrase with a full name phrase.

31. The method of generating a summary of claim 28, wherein the at least one rule to combine extracted sentences includes determining if two sentences having a common subject are proximate and whether at least one sentence is marked for reduction then removing the subject of the second sentence and combining with the first sentence using the connective "and."

32. A method of identifying correspondence between phrases in a sentence in a summary and phrases in the original document corresponding to the summary comprising:

establishing a plurality of heuristic rules for identifying a cut and paste summarization operation;

building a probability model based on the heuristic rules; and

calculating the best solution of the probability model to map a correspondence between the summary phrases and the original phrases.

33. The method of claim 32, wherein the probability model is a Hidden Markov Model.

34. The method of claim 33, wherein a Viterbi algorithm is employed to calculate the best solution.

35. A corpus for a summarization system comprising:

a plurality of documents;

a plurality of human generated summaries associated with the plurality of documents;

a sentence combination subcorpus; and

a sentence reduction subcorpus.

36. The corpus of claim 35, wherein the sentence combination subcorpus includes at least one mapping between a summary sentence and at least two original sentences containing phrases in the summary sentence.

37. The corpus of claim 35, wherein the sentence reduction subcorpus includes at least one sentence pair, each sentence pair having a summary sentence and a corresponding original sentence.

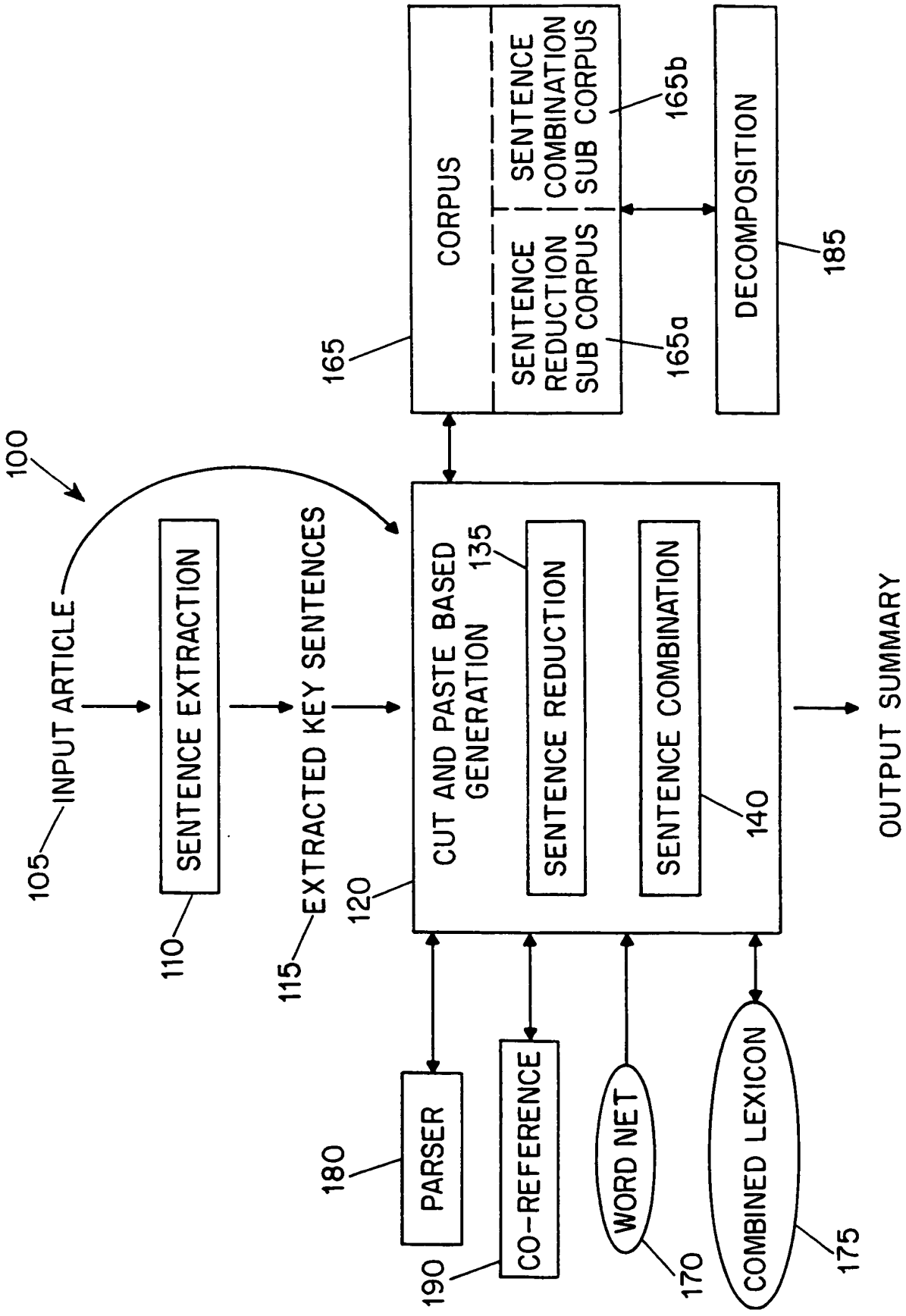
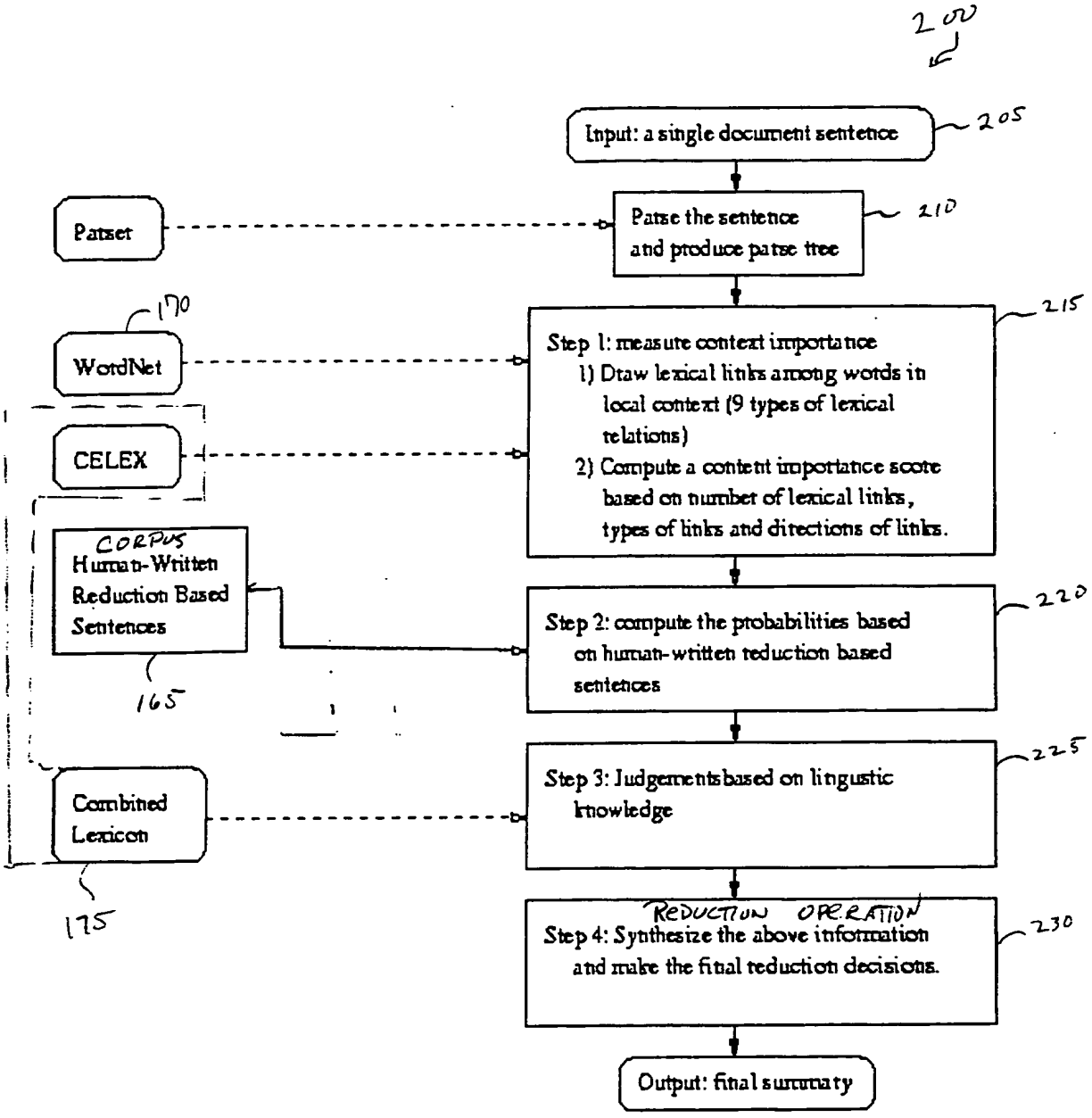

100
INPUT ARTICLE
105
SENTENCE EXTRACTION
110
EXTRACTED KEY SENTENCES
115
CUT AND PASTE BASED GENERATION
120
135
SENTENCE REDUCTION
SENTENCE COMBINATION
140
OUTPUT SUMMARY
CORPUS
165
SENTENCE REDUCTION SUB CORPUS
165a
SENTENCE COMBINATION SUB CORPUS
165b
DECOMPOSITION
185
PARSER
180
CO-REFERENCE
190
WORD NET
170
COMBINED LEXICON
175

FIG. 1

FIG. 2


200
Input: a single document sentence
205
Parse the sentence and produce parse tree
210
Parset
170
WordNet
CELEX
CORPUS
Human-Written Reduction Based Sentences
165
Step 1: measure context importance
1) Draw lexical links among words in local context (9 types of lexical relations)
2) Compute a content importance score based on number of lexical links, types of links and directions of links.
215
Step 2: compute the probabilities based on human-written reduction based sentences
220
Step 3: Judgements based on linguistic knowledge
225
Combined Lexicon
175
REDUCTION OPERATION
Step 4: Synthesize the above information and make the final reduction decisions.
230
Output: final summary

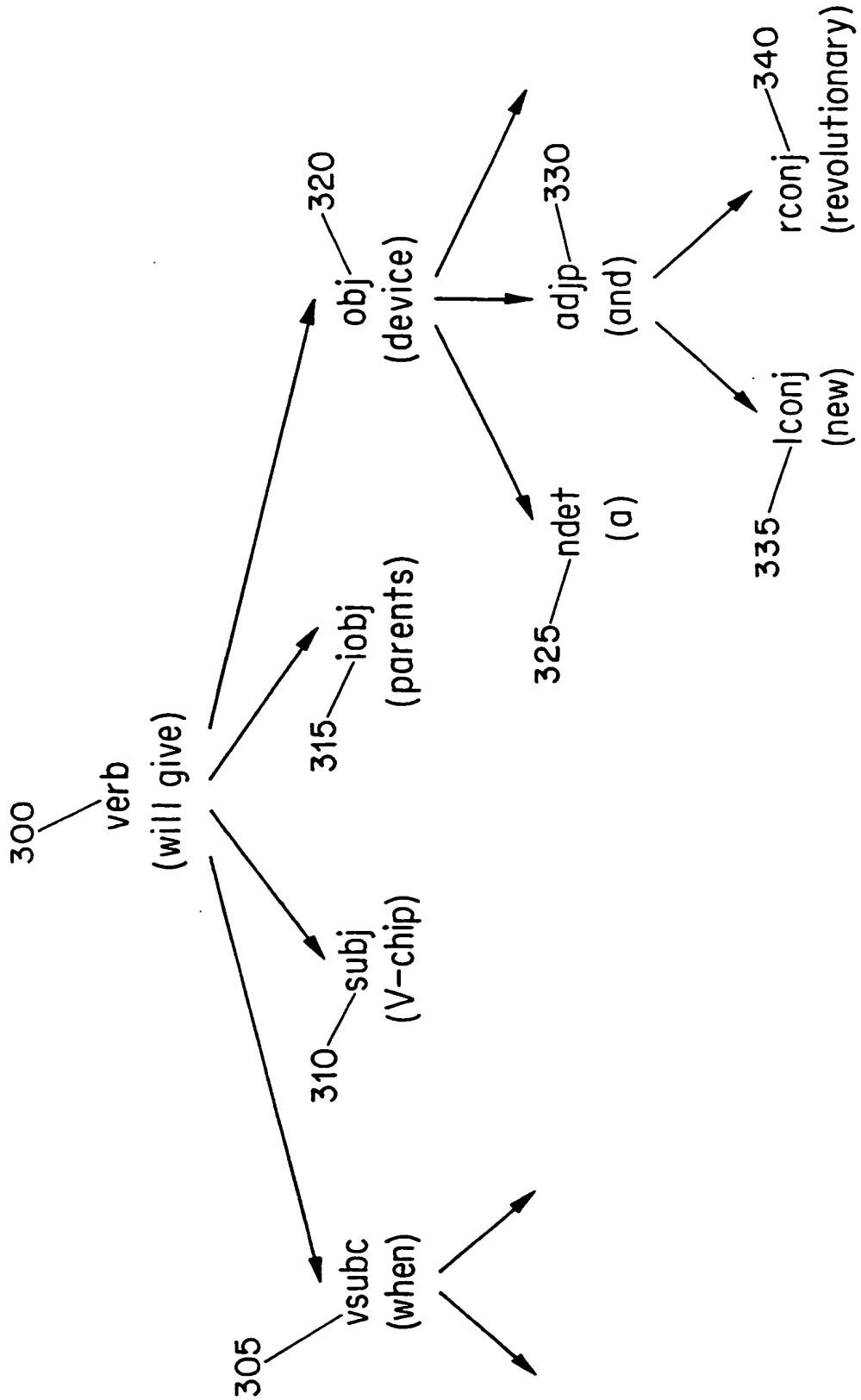

300
verb
(will give)
305
vsubc
(when)
310
subj
(V-chip)
315
iobj
(parents)
320
obj
(device)
325
ndet
(a)
330
adjp
(and)
335
lconj
(new)
340
rconj
(revolutionary)

FIG. 3

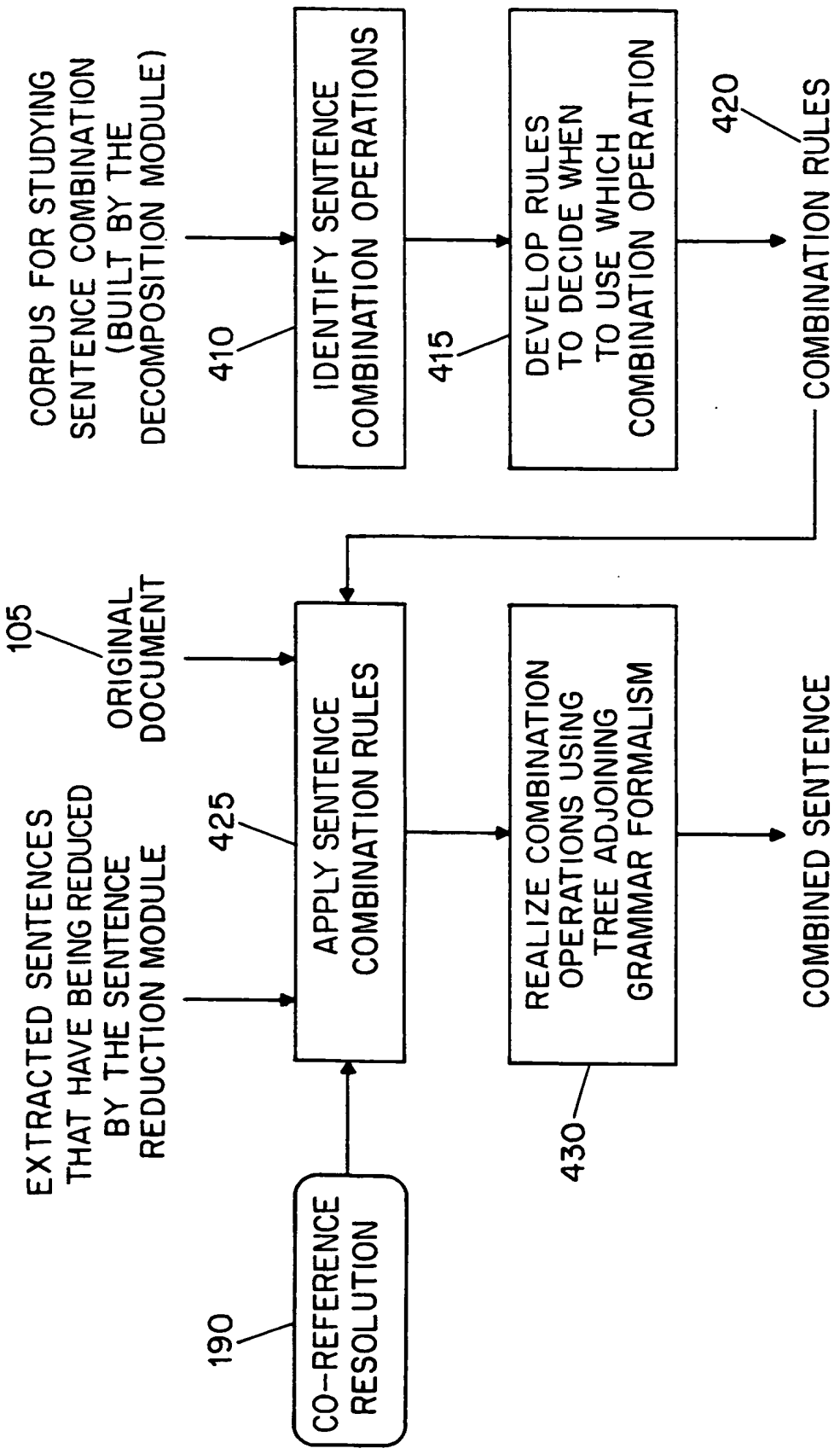

CORPUS FOR STUDYING SENTENCE COMBINATION (BUILT BY THE DECOMPOSITION MODULE)
410
IDENTIFY SENTENCE COMBINATION OPERATIONS
415
DEVELOP RULES TO DECIDE WHEN TO USE WHICH COMBINATION OPERATION
420
COMBINATION RULES
105
ORIGINAL DOCUMENT
EXTRACTED SENTENCES THAT HAVE BEING REDUCED BY THE SENTENCE REDUCTION MODULE
425
APPLY SENTENCE COMBINATION RULES
REALIZE COMBINATION OPERATIONS USING TREE ADJOINING GRAMMAR FORMALISM
430
COMBINED SENTENCE
190
CO-REFERENCE RESOLUTION

FIG. 4

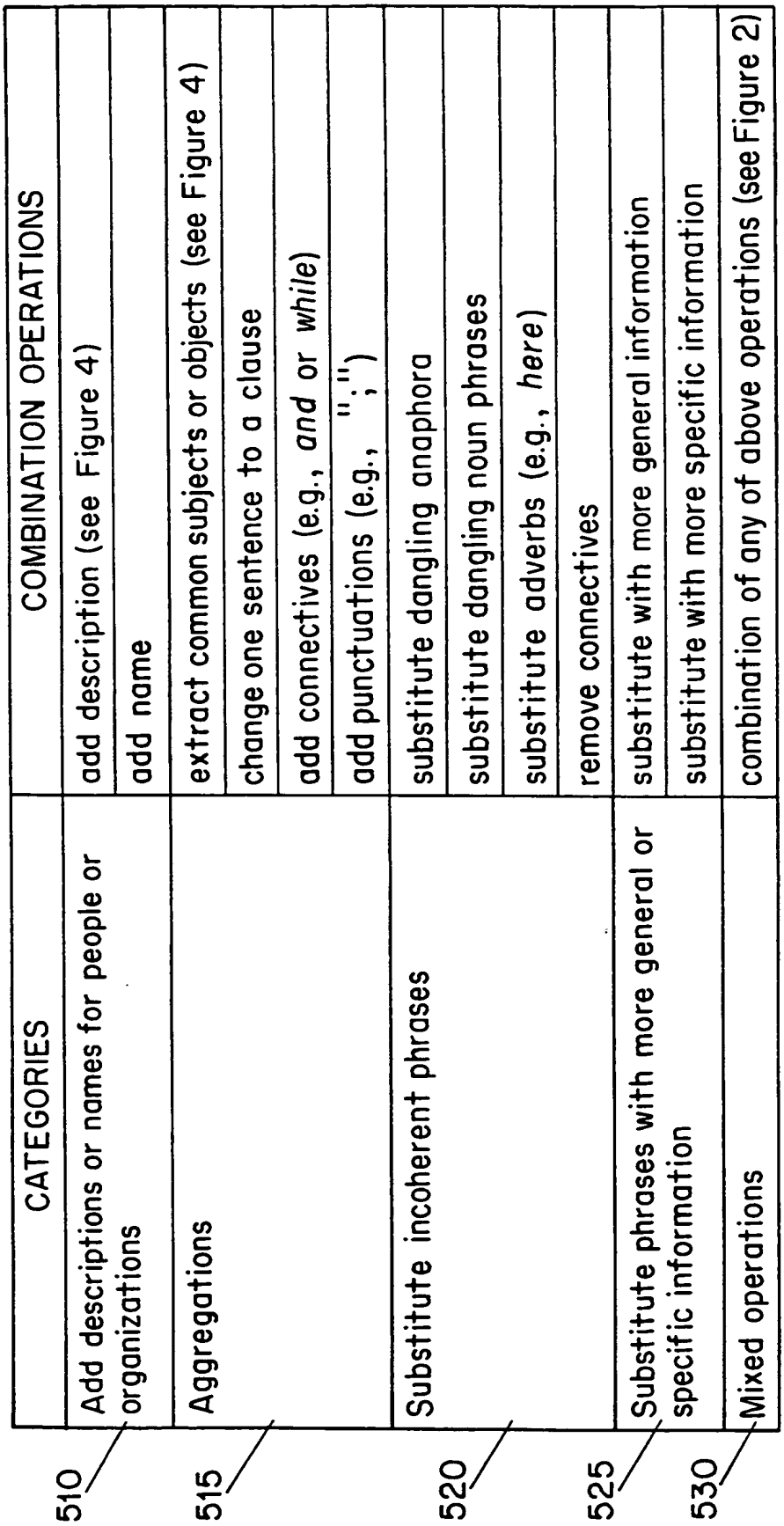

CATEGORIES
COMBINATION OPERATIONS
Add descriptions or names for people or organizations
add description (see Figure 4)
add name
Aggregations
extract common subjects or objects (see Figure 4)
change one sentence to a clause
add connectives (e.g., and or while)
add punctuations (e.g., ":" ";")
Substitute incoherent phrases
substitute dangling anaphora
substitute dangling noun phrases
substitute adverbs (e.g., here)
remove connectives
Substitute phrases with more general or specific information
substitute with more general information
substitute with more specific information
Mixed operations
combination of any of above operations (see Figure 2)
510
515
520
525
530

FIG. 5

RULE 1:

IF: ((a person or an organization is mentioned the first time) and (the full name or the full description of the person or the organization exists somewhere in the original article but is missing in the summary))

THEN: replace the phrase with the full name plus the full description

RULE 2:

IF: ((two sentences are close to each other in the original article) and (their subjects refer to the same entity) and (at least one of the sentences is the reduced form resulting from sentence reduction))

THEN: merge the two sentences by removing the subject in the second sentence, and then combining it with the first sentence using connective "and".

FIG. 6

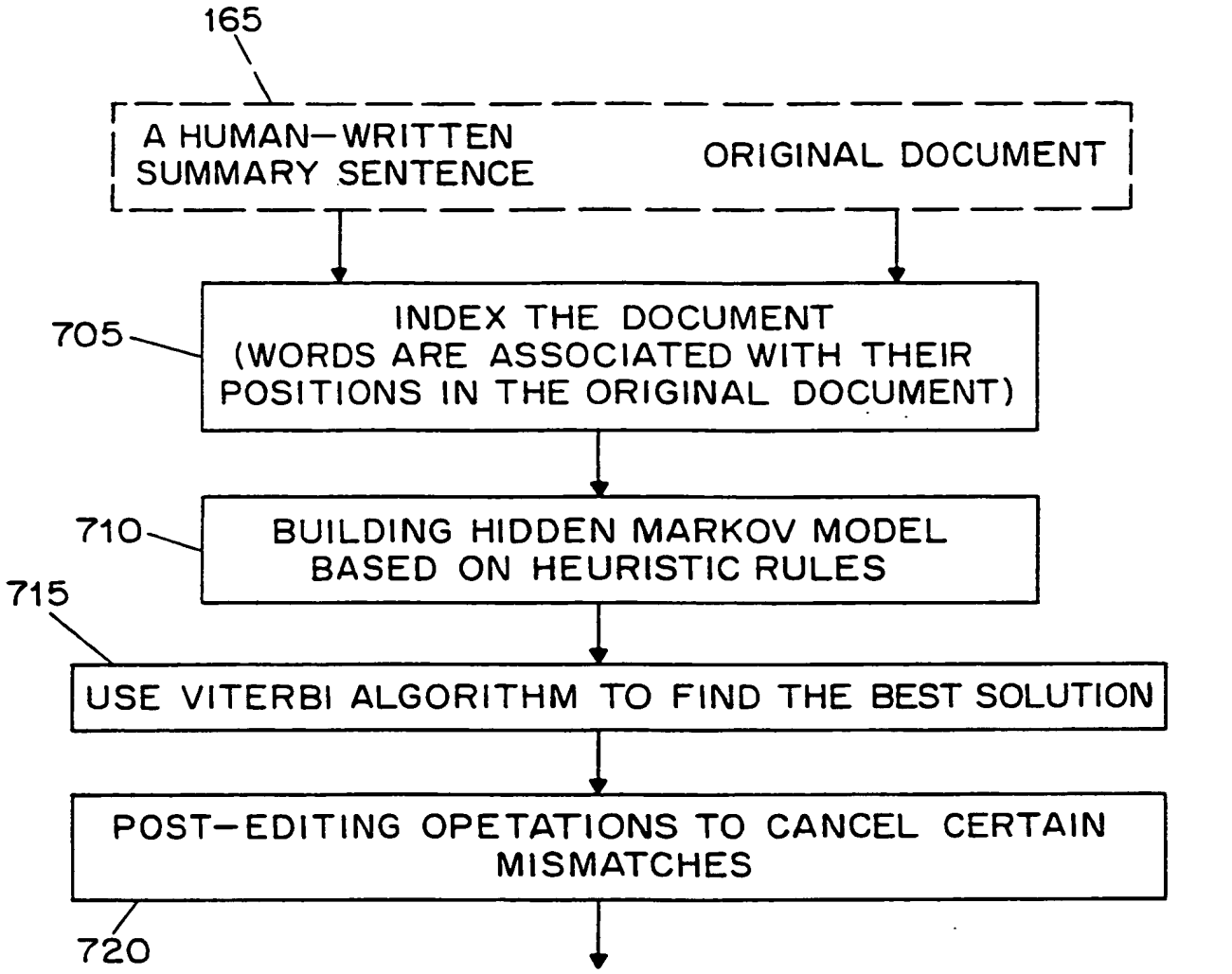

165
A HUMAN-WRITTEN SUMMARY SENTENCE
ORIGINAL DOCUMENT
705
INDEX THE DOCUMENT (WORDS ARE ASSOCIATED WITH THEIR POSITIONS IN THE ORIGINAL DOCUMENT)
710
BUILDING HIDDEN MARKOV MODEL BASED ON HEURISTIC RULES
715
USE VITERBI ALGORITHM TO FIND THE BEST SOLUTION
POST-EDITING OPETATIONS TO CANCEL CERTAIN MISMATCHES
720
PHRASES IN THE SUMMARY SENTENCE ARE MATCHED TO THEIR CORRESPONDING PHRASES IN THE DOCUMENT

FIG. 7

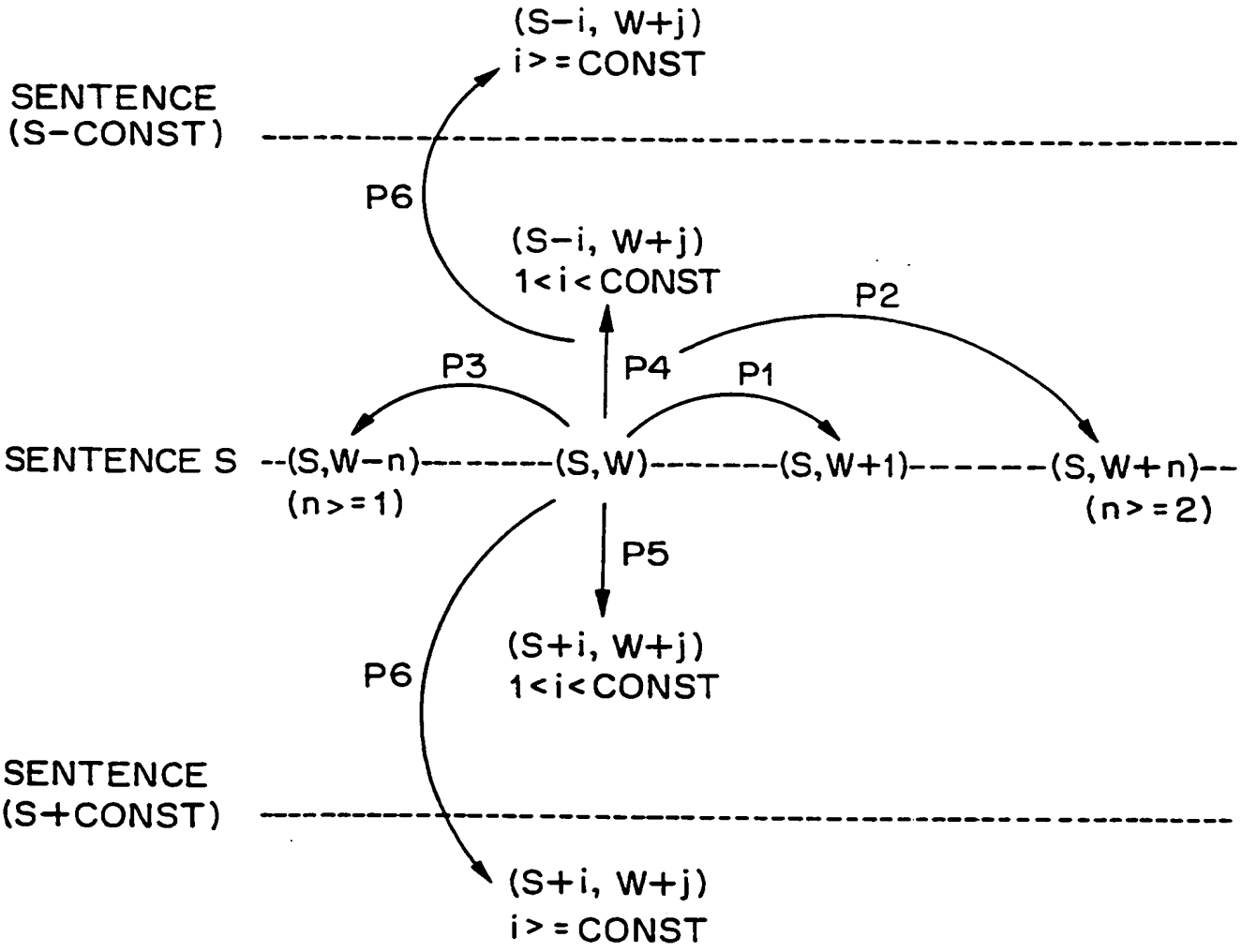

(S−i, W+j)
i>=CONST
SENTENCE
(S−CONST)
P6
(S−i, W+j)
1<i<CONST
P2
P3
P4
P1
SENTENCE S
--(S,W−n)--------(S,W)--------(S,W+1)--------(S,W+n)--
(n>=1)
(n>=2)
P5
(S+i, W+j)
1<i<CONST
P6
SENTENCE
(S+CONST)
(S+i, W+j)
i>=CONST

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04505

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :G06F 17/27
US CL : 707/500, 501, 530; 704/9, 10
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/500, 501, 530; 704/9, 10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST database
search terms: summary, summarization, document,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,778,397 A (KUPICE et al) 07 July 1998, col.3, line 37 to col.10, line 35	1-37
Y	US 5,077,668 A (DOI) 31 December 1991, col.2, line 50 to col.4, line 44.	1-37
A, P	US 5,918,240 A (KUPIEC et al) 29 June 1999, ALL	1-37
A	US 5,838,323 A (ROSE et al) 17 November 1998, ALL	1-37
A, P	US 5,924,108 A (FEIN et al) 13 July 1999, ALL	1-37

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents.	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
11 JULY 2000	15 AUG 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHEN HONG Telephone No. (703) 305-3900

Form PCT/ISA/210 (second sheet) (July 1998) *

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/04505

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 17/27

US CL : 707/500, 501, 530; 704/9, 10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/500, 501, 530; 704/9, 10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST database

search terms: summary, summarization, document,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,778,397 A (KUPICE et al) 07 July 1998, col.3, line 37 to col.10, line 35	1-37
Y	US 5,077,668 A (DOI) 31 December 1991, col.2, line 50 to col.4, line 44.	1-37
A, P	US 5,918,240 A (KUPIEC et al) 29 June 1999, ALL	1-37
A	US 5,838,323 A (ROSE et al) 17 November 1998, ALL	1-37
A, P	US 5,924,108 A (FEIN et al) 13 July 1999, ALL	1-37

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents.	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
11 JULY 2000	15 AUG 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHEN HONG Telephone No. (703) 305-3900

Form PCT/ISA/210 (second sheet) (July 1998) ★